```
                    GenCore version 5.1.6
             Copyright (c) 1993 - 2003  Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:         September 24, 2003, 20:25:50 ; Search time 6143 Seconds
                                             (without alignments)
                                             5503.423 Million cell updates/sec

Title:          US-09-512-581B-2
Perfect score:  7193
Sequence:       1 MAHSKTRTNDGKITYPPGVK..........QKGRGRPSKTPSPSQPKKNV 1391

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters:      45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09512581/runat_23092003_163547_5486/app_query.fasta_1.1543
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09512581_@CGN_1_1_4687_@runat_23092003_163547_5486 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      EST:*
                1:  em_estba:*
                2:  em_esthum:*
                3:  em_estin:*
                4:  em_estmu:*
                5:  em_estov:*
                6:  em_estpl:*
                7:  em_estro:*
                8:  em_htc:*
                9:  gb_est1:*
                10:  gb_est2:*
                11:  gb_htc:*
                12:  gb_est3:*
                13:  gb_est4:*
                14:  gb_est5:*
                15:  em_estfun:*
                16:  em_estom:*
                17:  em_gss_hum:*
                18:  em_gss_inv:*
                19:  em_gss_pln:*
                20:  em_gss_vrt:*
                21:  em_gss_fun:*
                22:  em_gss_mam:*
                23:  em_gss_mus:*
                24:  em_gss_pro:*
                25:  em_gss_rod:*
                26:  em_gss_phg:*
                27:  em_gss_vrl:*
                28:  gb_gss1:*
                29:  gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3787	52.6	3313	11	AK045159	AK045159 Mus muscu
2	3512	48.8	2899	11	AK041682	AK041682 Mus muscu
3	3309.5	46.0	4273	11	AK086753	AK086753 Mus muscu
4	3037	42.2	4427	11	AK032384	AK032384 Mus muscu
5	2170.5	30.2	1359	11	BC032988	BC032988 Mus muscu
6	1395	19.4	926	14	CA980171	CA980171 AGENCOURT
7	1325	18.4	786	14	CA324097	CA324097 UI-M-FY0-
8	1311	18.2	779	14	CA324284	CA324284 UI-M-FY0-
9	1298.5	18.1	786	13	BU703488	BU703488 UI-M-FO0-
10	1295	18.0	782	13	BU363024	BU363024 603790485
11	1285.5	17.9	943	13	BU914579	BU914579 AGENCOURT
12	1269.5	17.6	857	14	CA987660	CA987660 AGENCOURT
13	1262.5	17.6	882	13	BU916668	BU916668 AGENCOURT
14	1239	17.2	931	13	BU914025	BU914025 AGENCOURT
15	1231	17.1	829	13	BX314916	BX314916 BX314916
16	1225	17.0	791	10	BF509252	BF509252 UI-H-BI4-
17	1204.5	16.7	771	14	CB312535	CB312535 AGENCOURT
18	1189.5	16.5	908	13	BQ957734	BQ957734 AGENCOURT
19	1175.5	16.3	745	13	BU940978	BU940978 AGENCOURT
20	1168	16.2	692	12	BM794657	BM794657 K-EST0076
21	1162	16.2	716	14	CB522273	CB522273 UI-M-GH0-
22	1152.5	16.0	876	13	BU912768	BU912768 AGENCOURT
23	1143.5	15.9	734	10	BG256731	BG256731 602370939
24	1142	15.9	679	10	BB094813	BB094813 BB094813
25	1130	15.7	823	13	BU703952	BU703952 UI-M-FO0-
26	1128	15.7	703	14	CB521463	CB521463 UI-M-GH0-
27	1124.5	15.6	868	14	CD301710	CD301710 AGENCOURT
28	1119	15.6	658	14	CD542155	CD542155 B0238B04-
29	1108.5	15.4	695	14	CA405800	CA405800 1001949 H
30	1099	15.3	659	12	BI738240	BI738240 603361679
31	1091	15.2	729	10	BG432960	BG432960 602495783
32	1083	15.1	637	14	CB132602	CB132602 K-EST0183
33	1083	15.1	643	13	BU440284	BU440284 604146435
34	1078	15.0	874	13	BU517176	BU517176 AGENCOURT
35	1076	15.0	633	14	CD542034	CD542034 B0237B04-
36	1076	15.0	764	13	BU384740	BU384740 603860876
37	1071	14.9	773	9	AL705029	AL705029 DKFZp686B
38	1070	14.9	804	13	BU109808	BU109808 603128343
39	1069	14.9	638	14	CD284258	CD284258 G39173.86
40	1067.5	14.8	1142	12	BM466282	BM466282 AGENCOURT
41	1066	14.8	944	10	BF142489	BF142489 601788573
42	1061.5	14.8	848	10	BE873840	BE873840 601483992
43	1061	14.8	657	9	AL846882	AL846882 AL846882
44	1061	14.8	858	13	BQ736792	BQ736792 AGENCOURT
45	1056.5	14.7	624	14	CB215704	CB215704 NISC_np09

ALIGNMENTS

```
RESULT 1
AK045159
LOCUS       AK045159               3313 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
            enriched library, clone:B130042B12 product:49J10.1.2
            (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED
            PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo sapiens], full insert
            sequence.
ACCESSION   AK045159
VERSION     AK045159.1  GI:26337114
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3313)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. Tomohiro Kono (Department of Animal
            Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City,
            Kanagawa Prefecture, Japan) whose assistance we gratefully
            acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .3313
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:B130042B12"
                     /db_xref="taxon:10090"
                     /clone="B130042B12"
                     /tissue_type="parthenogenote"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="9.5 days embryo"
     CDS             <1. .2476
                     /note="unnamed protein product; 49J10.1.2
                     (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF
                     ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo
                     sapiens] (SPTR|Q96KV4, evidence: FASTY, 92.3%ID,
                     100%length, match=1430)
                     putative"
                     /codon_start=2
                     /protein_id="BAC32242.1"
                     /db_xref="GI:26337115"
                     /translation="VNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAE
                     TFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQA
                     KYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAP
                     LKSLVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNN
                     HSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIIT
                     LEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHA
                     RQCLVKNITVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKD
                     VKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDTKMNEKLYTVCDVAMN
                     IIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTANVLG
                     AVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMS
                     SPLPGKKSDKREDPDLVRSELEKPRSRKKAPVTDPEEKLGMDDLTKLVQEQKPKGSQR
                     GRKRGRTASDSDEQQWPEEKRHKEELLENEDEQNSPPKKGKRGRPPKPLGGGTSKEEP
                     TMKTSKKGNKKKLVPPVVDDDEEEERQIGNTEHKSKSKQHRTSKRAQQRAESPETSAV
                     ESTQSTPQKGRGRPSKAPSPSQPPKKIRVGRSKQVATKENDSSEEMDVLQASSPVSDD
                     TTQEGAEEEDISVGNVRRRSSKRERR"
     polyA_signal    3288. .3293
                     /note="putative"
     polyA_site      3313
                     /note="putative"
BASE COUNT     1066 a    709 c    732 g    806 t
ORIGIN

Alignment Scores:
Pred. No.:              7.51e-275     Length:         3313
Score:                  3787.00       Matches:        731
Percent Similarity:     97.00%        Conservative:   13
Best Local Similarity:  95.31%        Mismatches:     23
Query Match:            52.65%        Indels:         0
DB:                     11            Gaps:           0

US-09-512-581B-2 (1-1391) x AK045159 (1-3313)

Qy        625 ValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGln 644
```

```
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2 GTGAATAAATCAATAGATGGAACAGCAGATGATGAGGATGAGGGTGTTCCAACTGATCAA 61

Qy       645 AlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPhe 664
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        62 GCTATCAGGGCAGGCCTTGAGCTGCTGAAGGTGCTGTCTTTTACGCACCCTATCTCCTTT 121

Qy       665 HisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysVal 684
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       122 CACTCTGCTGAGACGTTTGAGTCCCTTCTGGCTTGTCTGAAGATGGATGATGAGAAGGTG 181

Qy       685 AlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPhePro 704
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       182 GCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAAGAGGATTTCCCT 241

Qy       705 HisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLysGlyProProArg 724
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       242 CACATCAGATCAGCCTTGCTTCCTGTACTACATCACAAATCCAAAAAAGGACCACCTCGG 301

Qy       725 GlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPhe 744
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       302 CAGGCCAAATACGCCATCCATTGTATTCATGCCATATTTTCTAGTAAAGAAACCCAGTTT 361

Qy       745 AlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIle 764
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       362 GCACAGATATTTGAGCCTCTGCATAAAAGCCTAGATCCAAGCAACCTGGAGCATCTTATA 421

Qy       765 ThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAlaPro 784
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       422 ACCCCTCTGGTCACTATTGGTCATATTGCTCTTCTCGCACCTGATCAGTTCGCTGCTCCT 481

Qy       785 TrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuPro 804
                |||||   |||||||||||||||||||||||||||||||||||||||||||||||||
Db       482 CTGAAGTCTTTGGTGGCAACTTTCATTGTGAAGGACCTCCTCATGAATGACCGGCTTCCA 541

Qy       805 GlyLysLysThrThrLysLeuTrpValProAspGluGluValSerProGluThrMetVal 824
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       542 GGAAAAAAGACAACTAAGCTTTGGGTTCCAGATGAGGAAGTCTCACCTGAGACAATGGTC 601

Qy       825 LysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnHisSer 844
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       602 AAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAGAATAATCACAGT 661

Qy       845 LysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeu 864
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       662 AAGTCAGGAACTTCCACCCTCAGATTACTAACAACGATACTGCATAGTGATGGGGATTTG 721

Qy       865 ThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlySer 884
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       722 ACAGAACAAGGGAAAATTAGTAAACCAGATATGTCACGCCTGAGACTTGCTGCTGGGAGT 781

Qy       885 AlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGluGlnTyr 904
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       782 GCTATTGTGAAGCTGGCACAGGAGCCCTGTTACCACGAGATCATTACACTGGAGCAGTAC 841

Qy       905 GlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGln 924
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       842 CAGCTGTGTGCATTAGCCATCAATGATGAGTGTTATCAAGTCAGGCAGGTGTTCGCTCAG 901

Qy       925 LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAla 944
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       902 AAACTTCACAAAGGCCTTTCCCGCTTACGGCTTCCCCTTGAGTACATGGCCATCTGTGCT 961

Qy       945 LeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLys 964
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       962 CTTTGTGCCAAAGACCCTGTGAAAGAGAGGCGAGCCCATGCTAGACAGTGCTTGGTGAAG 1021

Qy       965 AsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeu 984
             ||||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1022 AACATCACTGTGAGGAGGGAGTACCTGAAGCAGCATGCAGCTGTTAGTGAAAAATTATTG 1081

Qy       985 SerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAsp 1004
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1082 TCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCACATGACCCAGAT 1141

Qy      1005 TyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheVal 1024
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1142 TATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTGAAAGAATGCCTTTGGTTTGTT 1201

Qy      1025 LeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetVal 1044
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1202 CTGGAGATATTGATGGCTAAAAATGAAAACAACAGCCATGCATTTATCAGAAAAATGGTA 1261

Qy      1045 GluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLys 1064
             |||||||||||||||||||||||||||||||||||||||||  |||||||||||||||||
Db      1262 GAAAATATTAAACAGACAAAAGATGCTCAAGGACCAGATGATACAAAAATGAATGAAAAA 1321

Qy      1065 LeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSer 1084
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1322 TTGTACACCGTGTGTGATGTTGCCATGAACATCATCATGTCCAAGAGCACCACGTACAGC 1381

Qy      1085 LeuGluSerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsn 1104
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1382 CTGGAGTCTCCTAAGGACCCCGTGCTGCCAGCTCGGTTTTTCACCCAGCCTGACAAGAAT 1441

Qy      1105 PheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePheThrProGlyLys 1124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1442 TTTAGTAACACCAAAAATTACCTGCCTCCAGAAATGAAATCATTTTTCACTCCTGGAAAA 1501

Qy      1125 ProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGln 1144
             |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db      1502 CCTAAAACAGCCAATGTTCTCGGAGCAGTTAATAAGCCACTTTCATCAGCAGGCAAACAG 1561

Qy      1145 SerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnPro 1164
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1562 TCTCAGACCAAATCATCAAGAATGGAAACTGTGAGCAACGCAAGCAGCAGCTCCAACCCA 1621

Qy      1165 SerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspHisSerGluAsn 1184
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1622 AGCTCTCCTGGAAGGATCAAGGGGAGGCTTGATAGCTCTGAAATGGATCACAGTGAAAAT 1681

Qy      1185 GluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLysArgAspAspSer 1204
             ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db      1682 GAAGATTATACAATGTCTTCACCTTTGCCAGGAAAAAAAAGTGACAAGAGAGAAGACCCT 1741

Qy      1205 AspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThrProValThrGlu 1224
             ||||||||||||||||||||||||||||||||   |||||||||   |||||||||:::
Db      1742 GATCTTGTAAGGTCTGAGTTGGAGAAGCCTAGAAGTCGGAAAAAAGCACCTGTCACAGAC 1801

Qy      1225 GlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGluGlnLysProLys 1244
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1802 CCTGAAGAGAAATTAGGTATGGATGACCTAACTAAGTTGGTACAGGAACAGAAACCTAAA 1861

Qy      1245 GlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrp 1264
             |||||||||||   |||||||||||   ||||||||||:::||||||||||||||||||
Db      1862 GGCAGTCAGCGAGGCCGGAAAAGAGGCCGTACAGCTTCAGACTCAGACGAGCAGCAGTGG 1921

Qy      1265 ProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsnSerPro 1284
             ||||||||||||||   ||||||::::::||||||||||||||||||||||||||||||
Db      1922 CCTGAGGAGAAGAGGCACAAAGAGGAGCTCCTGGAAAATGAGGATGAGCAGAACAGCCCA 1981

Qy      1285 ProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGlyThrProLysGlu 1304
             ||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db      1982 CCAAAAAAGGGCAAAAGAGGCAGGCCACCAAAACCTCTTGGTGGGGGACATCGAAGGAA 2041

Qy      1305 GluProThrMetLysThrSerLysLysGlySerLysLysLysSerGlyProProAlaPro 1324
             ||||||||||||||||||||||||||||||||:::||||||||||      |||||
Db      2042 GAGCCCACAATGAAAACATCCAAGAAAGGAAACAAGAAAAAACTTGTACCTCCTGTAGTA 2101

Qy      1325 GluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSerLys 1344
             :::::::::||||||||||||||||||   ||||||||||||   |||||||||||||||
Db      2102 GACGACGATGAAGAAGAAGAAAGACAAATTGGAAACACAGAACATAAGTCAAAAAGCAAA 2161

Qy      1345 GlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGluSerSerAlaIle 1364
             |||||||||   |||:::||||||||||||||||||||||||||||||:::||||||:::
Db      2162 CAGCACCGCACATCAAAGAGAGCACAACAGAGAGCAGAATCTCCTGAAACAAGTGCAGTT 2221
```

```
Qy        1365 GluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLysThrProSerPro 1384
               ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        2222 GAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAAGCACCATCACCA 2281

Qy        1385 SerGlnProLysLysAsnVal 1391
               |||||||||   |||   :::
Db        2282 TCACAACCCCCCAAAAAAATC 2302

RESULT 2
AK041682
LOCUS       AK041682                2899 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
            library, clone:A630029M15 product:49J10.1.2 (ANDROGEN-INDUCED
            PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2)
            (FRAGMENT) homolog [Homo sapiens], full insert sequence.
ACCESSION   AK041682
VERSION     AK041682.1  GI:26334660
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2899)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
            Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
            Trust/MRC building Addenbrookes Hospital Cambridge) whose
            assistance we gratefully acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..2899
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:A630029M15"
                     /db_xref="taxon:10090"
                     /clone="A630029M15"
                     /tissue_type="thymus"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="3 days neonate"
     CDS             117..2222
                     /note="unnamed protein product; 49J10.1.2
                     (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF
                     ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo
                     sapiens] (SPTR|Q96KV4, evidence: FASTY, 92.3%ID,
                     100%length, match=1430)
                     putative"
                     /codon_start=1
                     /protein_id="BAC31031.1"
                     /db_xref="GI:26334661"
                     /translation="MVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEALVSPTCSC
                     KQAEGCVREITKKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQVNKSIDG
                     TADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAAL
                     QIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQI
                     FEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPLKSLVATFIVKDLLMNDRLPG
                     KKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGD
                     LTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQV
                     FAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNITVRREYLKQHAAV
                     SEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSH
                     AFIRKMVENIKQTKDAQGPDDTKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPA
                     RFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTANVLGAVNKPLSSAGKQSQTKSSRME
                     TVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKREDPDLVRVRC
```

```
                    LVGRVMRLLIVIVLVIFAL"
BASE COUNT        868 a     582 c     617 g     832 t
ORIGIN

Alignment Scores:
Pred. No.:              3.51e-254      Length:        2899
Score:                  3512.00        Matches:       683
Percent Similarity:     98.99%         Conservative:  2
Best Local Similarity:  98.70%         Mismatches:    7
Query Match:            48.83%         Indels:        0
DB:                     11             Gaps:          0

US-09-512-581B-2 (1-1391) x AK041682 (1-2899)

Qy        517 LysThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeu 536
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 CAGACAGATGCCAGTGTCAAGGCCATATTTTCTAAAGTGATGGTTATCACAAGAAATTTG 137

Qy        537 ProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAsp 556
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 CCAGATCCTGGTAAGGCCCAAGACTTCATGAAGAAGTTCACACAAGTGTTGGAAGATGAT 197

Qy        557 GluLysIleArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAla 576
              ||||||||||||||||||||||||   |||||||||||||||||||||||||||||||||
Db        198 GAGAAAATACGGAAACAGCTGGAAGCCCTTGTTAGTCCCACATGTTCCTGCAAGCAGGCT 257

Qy        577 GluGlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnPro 596
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 GAAGGCTGTGTGCGTGAAATAACTAAGAAGTTGGGCAACCCTAAGCAGCCTACAAACCCT 317

Qy        597 PheLeuGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGlu 616
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 TTCTTGGAAATGATCAAGTTTCTCTTGGAGAGGATAGCTCCTGTACACATAGACACTGAA 377

Qy        617 SerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGlu 636
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 TCCATCAGTGCTCTTATCAAGCAAGTGAATAAATCAATAGATGGAACAGCAGATGATGAG 437

Qy        637 AspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeu 656
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 GATGAGGGTGTTCCAACTGATCAAGCTATCAGGGCAGGCCTTGAGCTGCTGAAGGTGCTG 497

Qy        657 SerPheThrHisProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCys 676
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 TCTTTTACGCACCCTATCTCCTTTCACTCTGCTGAGACGTTTGAGTCCCTTCTGGCTTGT 557

Qy        677 LeuLysMetAspAspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGly 696
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 CTGAAGATGGATGATGAGAAGGTGGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGA 617

Qy        697 SerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHis 716
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 AGCAAAATTGAAGAGGATTTCCCTCACATCAGATCAGCCTTGCTTCCTGTACTACATCAC 677

Qy        717 LysSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIle 736
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        678 AAATCCAAAAAAGGACCACCTCGGCAGGCCAAATACGCCATCCATTGTATTCATGCCATA 737

Qy        737 PheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAsp 756
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        738 TTTTCTAGTAAAGAAACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAAAGCCTAGAT 797

Qy        757 ProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeu 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        798 CCAAGCAACCTGGAGCATCTTATAACCCCTCTGGTCACTATTGGTCATATTGCTCTTCTC 857

Qy        777 AlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAsp 796
              ||||||||||||||||||||||||    ||||||    ||||||||||||||||||||||
Db        858 GCACCTGATCAGTTCGCTGCTCCTCTGAAGTCTTTGGTGGCAACTTTCATTGTGAAGGAC 917

Qy        797 LeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGlu 816
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 CTCCTCATGAATGACCGGCTTCCAGGAAAAAAGACAACTAAGCTTTGGGTTCCAGATGAG 977

Qy        817 GluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeu 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 GAAGTCTCACCTGAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTA 1037

Qy        837 LeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThr 856
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 CTTGGAATGAAGAATAATCACAGTAAGTCAGGAACTTCCACCCTCAGATTACTAACAACG 1097

Qy        857 IleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSer 876
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1098 ATACTGCATAGTGATGGGGATTTGACAGAACAAGGGAAAATTAGTAAACCAGATATGTCA 1157

Qy        877 ArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHis 896
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1158 CGCCTGAGACTTGCTGCTGGGAGTGCTATTGTGAAGCTGGCACAGGAGCCCTGTTACCAC 1217

Qy        897 GluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyr 916
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1218 GAGATCATTACACTGGAGCAGTACCAGCTGTGTGCATTAGCCATCAATGATGAGTGTTAT 1277

Qy        917 GlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuPro 936
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1278 CAAGTCAGGCAGGTGTTCGCTCAGAAACTTCACAAAGGCCTTTCCCGCTTACGGCTTCCC 1337

Qy        937 LeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAla 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 CTTGAGTACATGGCCATCTGTGCTCTTTGTGCCAAAGACCCTGTGAAAGAGAGGCGAGCC 1397

Qy        957 HisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHis 976
              |||||||||||||||||||||||||||||    |||||||||||||||||||||||||||
Db       1398 CATGCTAGACAGTGCTTGGTGAAGAACATCACTGTGAGGAGGGAGTACCTGAAGCAGCAT 1457

Qy        977 AlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIle 996
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1458 GCAGCTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATT 1517

Qy        997 HisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAsp 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1518 CACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGAT 1577

Qy       1017 ValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSer 1036
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1578 GTGAAAGAATGCCTTTGGTTTGTTCTGGAGATATTGATGGCTAAAAATGAAAACAACAGC 1637

Qy       1037 HisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPro 1056
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1638 CATGCATTTATCAGAAAAATGGTAGAAAATATTAAACAGACAAAAGATGCTCAAGGACCA 1697

Qy       1057 AspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIle 1076
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1698 GATGATACAAAAATGAATGAAAAATTGTACACCGTGTGTGATGTTGCCATGAACATCATC 1757

Qy       1077 MetSerLysSerThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArg 1096
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1758 ATGTCCAAGAGCACCACGTACAGCCTGGAGTCTCCTAAGGACCCCGTGCTGCCAGCTCGG 1817

Qy       1097 PhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMet 1116
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1818 TTTTTCACCCAGCCTGACAAGAATTTTAGTAACACCAAAAATTACCTGCCTCCAGAAATG 1877

Qy       1117 LysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLys 1136
              ||||||||||||||||||||||||||||||     |||||||||||||||||||||||||
Db       1878 AAATCATTTTTCACTCCTGGAAAACCTAAAACAGCCAATGTTCTCGGAGCAGTTAATAAG 1937

Qy       1137 ProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSer 1156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1938 CCACTTTCATCAGCAGGCAAACAGTCTCAGACCAAATCATCAAGAATGGAAACTGTGAGC 1997

Qy       1157 AsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSer 1176
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1998 AACGCAAGCAGCAGCTCCAACCCAAGCTCTCCTGGAAGGATCAAGGGGAGGCTTGATAGC 2057
```

```
Qy       1177 SerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLys 1196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2058 TCTGAAATGGATCACAGTGAAAATGAAGATTATACAATGTCTTCACCTTTGCCAGGAAAA 2117

Qy       1197 LysSerAspLysArgAspAspSerAspLeuValArg 1208
              |||||||||||||||:::|||   ||||||||||||
Db       2118 AAAAGTGACAAGAGAGAAGACCCTGATCTTGTAAGG 2153

RESULT 3
AK086753
LOCUS       AK086753                4273 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 15 days embryo head cDNA, RIKEN full-length enriched
            library, clone:D930049H23 product:49J10.1.2 (ANDROGEN-INDUCED
            PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2)
            (FRAGMENT) homolog [Homo sapiens], full insert sequence.
ACCESSION   AK086753
VERSION     AK086753.1  GI:26103739
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 4273)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..4273
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:D930049H23"
                     /db_xref="taxon:10090"
                     /clone="D930049H23"
                     /tissue_type="head"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="15 days embryo"
     misc_feature    <1..2245
                     /note="49J10.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE
                     SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog
                     [Homo sapiens] (SPTR|Q96KV4, evidence: FASTY, 92.3%ID,
                     100%length, match=1430)
                     putative"
BASE COUNT     1363 a    840 c    899 g   1171 t
ORIGIN

Alignment Scores:
Pred. No.:              1.06e-238     Length:        4273
Score:                  3309.50       Matches:       652
Percent Similarity:     96.24%        Conservative:  14
Best Local Similarity:  94.22%        Mismatches:    23
Query Match:            46.01%        Indels:        5
DB:                     11            Gaps:          1

US-09-512-581B-2 (1-1391) x AK086753 (1-4273)

Qy        702 AspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSer-LysLysGl 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 GATTTCCCTCACATCAGATCAGCCTTGCTTCCTGTACTACATCACAAATCCAAAAAAAGG 60

Qy        721 yProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGl 741
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACCACCTCGGCAGGCCAAATACGCCATCCATTGTATTCATGCCATATTTTCTAGTAAAGA 120
```

```
Qy        741 uThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeuGl 761
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAAAGCCTAGATCCAAGCAACCTGGA 180

Qy        761 uHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPh 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCATCTTATAACCCCTCTGGTCACTATTGGTCATATTGCTCTTCTCGCACCTGATCAGTT 240

Qy        781 eAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsnAs 801
              ||||||||||   ||||||  |||||||||||||||||||||||||||||||||||||||
Db        241 CGCTGCTCCTCTGAAGTCTTTGGTGGCAACTTTCATTGTGAAGGACCTCCTCATGAATGA 300

Qy        801 pArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerProGl 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCGGCTTCCAGGAAAAAAGACAACTAAGCTTTGGGTTCCAGATGAGGAAGTCTCACCTGA 360

Qy        821 uThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAs 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAGAA 420

Qy        841 nAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAs 861
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TAATCACAGTAAGTCAGGAACTTCCACCCTCAGATTACTAACAACGATACTGCATAGTGA 480

Qy        861 pGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAl 881
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TGGGGATTTGACAGAACAAGGGAAAATTAGTAAACCAGATATGTCACGCCTGAGACTTGC 540

Qy        881 aAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThrLe 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGCTGGGAGTGCTATTGTGAAGCTGGCACAGGAGCCCTGTTACCACGAGATCATTACACT 600

Qy        901 uGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnVa 921
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db        601 GGAGCAGTACCAGCTGTGTGCATTAGCCATCAATGATGAGTGTTATCAAGTCAGGCAGGT 660

Qy        921 lPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeu-ProLeuGluTyrMetA 941
              ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db        661 GTTCGCTCAGAAACTTCACAAAGGCCTTTCCCGCTTACGGCTTCCCCTTGGAGTACATGG 720

Qy        941 laIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnC 961
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||||
Db        721 CCATCTGTGCTCTTTGTGCCAAAGACCCTGTGAAAGAGAGGGGAGCCCATGCTAGACAGT 780

Qy        961 ysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerG 981
              ||||||||||||||||    ||||||||||||||||||||||||||||||||||||||||
Db        781 GCTTGGTGAAGAACATCCC-GTGAGGAGGGAGTACCTGAAGCAGCATGCAGCTGTTAGTG 839

Qy        981 luLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaH 1001
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 AAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCAC 899

Qy       1001 isAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysL 1021
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 ATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTGAAAGAATGCC 959

Qy       1021 euTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleA 1041
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 TTTGGTTTGTTCTGGAGATATTGATGGCTAAAAATGAAAACAACAGCCATGCATTTATCA 1019

Qy       1041 rgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysM 1061
              |||||||||||||||||||||||||||||||||||||||||||||||||||||   ||||
Db       1020 GAAAAATGGTAGAAAATATTAAACAGACAAAAGATGCTCAAGGACCAGATGATACAAAAA 1079

Qy       1061 etAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSerT 1081
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 TGAATGAAAAATTGTACACCGTGTGTGATGTTGCCATGAACATCATCATGTCCAAGAGCA 1139

Qy       1081 hrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGlnP 1101
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1140 CCACGTACAGCCTGGAGTCTCCTAAGGACCCCGTGCTGCCAGCTCGGTTTTTCACCCAGC 1199

Qy       1101 roAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePheT 1121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1200 CTGACAAGAATTTTAGTAACACCAAAAATTACCTGCCTCCAGAAATGAAATCATTTTTCA 1259

Qy       1121 hrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerA 1141
              ||||||||||||||||||    ||||||||||||||||||||||||||||||||||||||
Db       1260 CTCCTGGAAAACCTAAAACAGCCAATGTTCTCGGAGCAGTTAATAAGCCACTTTCATCAG 1319

Qy       1141 laGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSerS 1161
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1320 CAGGCAAACAGTCTCAGACCAAATCATCAAGAATGGAAACTGTGAGCAACGCAAGCAGCA 1379

Qy       1161 erSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspH 1181
              ||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db       1380 GCTCCAACCCAAGCTCTCCTGGAAGGATCAG-GGGAGGCTTGATAGCTCTGAAATGGATC 1438

Qy       1181 isSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLysA 1201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 ACAGTGAAAATGAAGATTATACAATGTCTTCACCTTTGCCAGGAAAAAAAAGTGACAAGA 1498

Qy       1201 rgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThrP 1221
              ||:::|||   ||||||||||||||||||||||||||||||   |||||||||   |
Db       1499 GAGAAGACCCTGATCTTGTAAGGTCTGAGTTGGAGAAGCCTAGAAGTCGGAAAAAAGCAC 1558

Qy       1221 roValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGluG 1241
              ||||||||:::    |||||||||||||||||||||||||||||||||||||||||||||
Db       1559 CTGTCACAGACCCTGAAGAGAAATTAGGTATGGATGACCTAACTAAGTTGGTACAGGAAC 1618

Qy       1241 lnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspG 1261
              |||||||||||||||||||||   ||||||||||||   |||||||||:::|||||||||
Db       1619 AGAAACCTAAAGGCAGTCAGCGAGGCCGGAAAAGAGGCCGTACAGCTTCAGACTCAGACG 1678

Qy       1261 luGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluG 1281
              ||||||||||||||||||||||||    ||||||:::::|||||||||||||||||||||
Db       1679 AGCAGCAGTGGCCTGAGGAGAAGAGGCACAAAGAGGAGCTCCTGGAAAATGAGGATGAGC 1738

Qy       1281 lnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGlyT 1301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1739 AGAACAGCCCACCAAAAAAGGGCAAAAGAGGCAGGCCACCCAAACCTCTTGGTGGGGGGA 1798

Qy       1301 hrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLysSerGlyP 1321
              ||    ||||||||||||||||||||||||||||||||||:::||||||||||    |
Db       1799 CATCGAAGGAAGAGCCCACAATGAAAACATCCAAGAAAGGAAACAAGAAAAAACTTGTAC 1858

Qy       1321 roProAlaProGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLysS 1341
              |||||      :::::::::||||||:::||||||||||   |||||||||||   |||
Db       1859 CTCCTGTAGTAGACGACGATGAAGAACAAGAAAGACAAATTGGAAACACAGAACATAAGT 1918

Qy       1341 erLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGluS 1361
              ||||||||||||||||||    |||:::||||||||||||||||||||||||||||||:
Db       1919 CAAAAAGCAAACAGCACCGCACATCAAAGAGAGCACAACAGAGAGCAGAATCTCCTGAAA 1978

Qy       1361 erSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLysT 1381
              ::||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db       1979 CAAGTGCAGTTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAAG 2038

Qy       1381 hrProSerProSerGln---ProLysLysAsn 1390
              ||||||||||||||   |||:::||||||
Db       2039 CACCATCACCATCACAACCCCCCCAAAAAAAT 2070
```

RESULT 4
AK032384

```
LOCUS       AK032384               4427 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus adult male olfactory brain cDNA, RIKEN full-length
            enriched library, clone:6430530P15 product:49J10.1.2
            (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED
            PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo sapiens], full insert
            sequence.
ACCESSION   AK032384
VERSION     AK032384.1  GI:26082814
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 4427)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .4427
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:6430530P15"
                     /db_xref="taxon:10090"
                     /clone="6430530P15"
                     /sex="male"
                     /tissue_type="olfactory brain"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     misc_feature    <1. .2062
                     /note="49J10.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE
                     SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog
                     [Homo sapiens] (SPTR|Q96KV4, evidence: FASTY, 92.3%ID,
                     100%length, match=1430)
                     putative"
     polyA_signal    4406. .4411
                     /note="putative"
     polyA_site      4427
                     /note="putative"
BASE COUNT     1423 a    832 c    919 g   1253 t
ORIGIN

Alignment Scores:
Pred. No.:              3.96e-218       Length:         4427
Score:                  3037.00         Matches:        591
Percent Similarity:     95.88%          Conservative:   14
Best Local Similarity:  93.66%          Mismatches:     24
Query Match:            42.22%          Indels:         3
DB:                     11              Gaps:           1

US-09-512-581B-2 (1-1391) x AK032384 (1-4427)

Qy        761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 GAGCATCTTATAACCCCTCTGGTCACTATTGGTCATATTGCTCTTCTCGCACCTGATCAG 62

Qy        781 PheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
              ||||||||||||    ||||||    ||||||||||||||||||||||||||||||||||
Db         63 TTCGCTGCTCCTCTGAAGTCTTTGGTGGCAACTTTCATTGTGAAGGACCTCCTCATGAAT 122

Qy        801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerPro 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 GACCGGCTTCCAGGAAAAAAGACAACTAAGCTTTGGGTTCCAGATGAGGAAGTCTCACCT 182

Qy        821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLys 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAG 242

Qy        841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 AATAATCACAGTAAGTCAGGAACTTCCACCCTCAGATTACTAACAACGATACTGCATAGT 302

Qy        861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 GATGGGGATTTGACAGAACAAGGGAAAATTAGTAAACCAGATATGTCACGCCTGAGACTT 362
```

```
Qy        881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAGGAGCCCTGTTACCACGAGATCATTACA 422

Qy        901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 CTGGAGCAGTACCAGCTGTGTGCATTAGCCATCAATGATGAGTGTTATCAAGTCAGGCAG 482

Qy        921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet 940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 GTGTTCGCTCAGAAACTTCACAAAGGCCTTTCCCGCTTACGGCTTCCCCTTGAGTACATG 542

Qy        941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGln 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 GCCATCTGTGCTCTTTGTGCCAAAGACCCTGTGAAAGAGAGGCGAGCCCATGCTAGACAG 602

Qy        961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSer 980
              ||||||||||||||||    ||||||||||||||||||||||||||||||||||||||||
Db        603 TGCTTGGTGAAGAACATCACTGTGAGGAGGGAGTACCTGAAGCAGCATGCAGCTGTTAGT 662

Qy        981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
              ||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 GAAAAATTAATGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 722

Qy       1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        723 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTGAAAGAATGC 782

Qy       1021 LeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle 1040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        783 CTTTGGTTTGTTCTGGAGATATTGATGGCTAAAAATGAAAACAACAGCCATGCATTTATC 842

Qy       1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        843 AGAAAAATGGTAGAAAATATTAAACAGACAAAAGATGCTCAAGGACCAGATGATACAAAA 902

Qy       1061 MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        903 ATGAATGAAAAATTGTACACCGTGTGTGATGTTGCCATGAACATCATCATGTCCAAGAGC 962

Qy       1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln 1100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        963 ACCACGTACAGCCTGGAGTCTCCTAAGGACCCCGTGCTGCCAGCTCGGTTTTTCACCCAG 1022

Qy       1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe 1120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1023 CCTGACAAGAATTTTAGTAACACCAAAAATTACCTGCCTCCAGAAATGAAATCATTTTTC 1082

Qy       1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
              |||||||||||||||||||    |||||||||||||||    ||||||||||||||||||
Db       1083 ACTCCTGGAAAACCTAAAACAGCCAATGTTCTCGGAGCAGT-AATAAGCCACTTTCATCA 1141

Qy       1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1142 GCAGGCAAACAGTCTCAGACCAAATCATCAAGAATGGAAACTGTGAGCAACGCAAGCAGC 1201

Qy       1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1202 AGCTCCAACCCAAGCTCTCCTGGAAGGATCAAGGGGAGGCTTGATAGCTCTGAAATGGAT 1261

Qy       1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1262 CACAGTGAAAATGAAGATTATACAATGTCTTCACCTTTGCCAGGAAAAAAAAGTGACAAG 1321

Qy       1201 ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThr 1220
              |||:::|||  ||||||        ||||||||||||||||||||   |||||||||
Db       1322 AGAGAAGACCCTGATCTT------TCTGAGTTGGAGAAGCCTAGAAGTCGGAAAAAAGCA 1375

Qy       1221 ProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlu 1240
              |||||||||:::  ||||||||||||||||||||||||||||||||||||||||||||||
Db       1376 CCTGTCACAGACCCTGAAGAGAAATTAGGTATGGATGACCTAACTAAGTTGGTACAGGAA 1435

Qy       1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp 1260
              ||||||||||||||||||||||||     ||||||||||||    |||||||||:::||||||
Db       1436 CAGAAACCTAAAGGCAGTCAGCGAGGCCGGAAAAGAGGCCGTACAGCTTCAGACTCAGAC 1495

Qy       1261 GluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGlu 1280
              ||||||||||||||||||||||||     |||||:::::||||||||||||||||||
Db       1496 GAGCAGCAGTGGCCTGAGGAGAAGAGGCACAAAGAGGAGCTCCTGGAAAATGAGGATGAG 1555

Qy       1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1556 CAGAACAGCCCACCAAAAAAGGGCAAAAGAGGCAGGCCACCAAAACCTCTTGGTGGGGGG 1615

Qy       1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLysSerGly 1320
              |||   |||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db       1616 ACATCGAAGGAAGAGCCCACAATGAAAACATCCAAGAAAGGAAACAAGAAAAAAACTTGTA 1675

Qy       1321 ProProAlaProGluGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys 1340
              ||||||      :::::::::||||||||||||||||   |||||||||||   |||
Db       1676 CCTCCTGTAGTAGACGACGATGAAGAAGAAGAAAGACAAATTGGAAACACAGAACATAAG 1735

Qy       1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
              ||||||||||||||||||||     |||:::||||||||||||||||||||||||||||||
Db       1736 TCAAAAAGCAAACAGCACCGCACATCAAAGAGAGCACAACAGAGAGCAGAATCTCCTGAA 1795

Qy       1361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
              :::||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db       1796 ACAAGTGCAGTTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 1855

Qy       1381 ThrProSerProSerGlnProLysLysAsnVal 1391
                 ||||||||||||||||   |||   :::
Db       1856 GCACCATCACCATCACAACCCCCCAAAAAAATC 1888
```

RESULT 5
BC032988

```
LOCUS       BC032988                1359 bp    mRNA    linear   HTC 20-SEP-2002
DEFINITION  Mus musculus, clone IMAGE:1179696, mRNA.
ACCESSION   BC032988
VERSION     BC032988.1  GI:21426933
KEYWORDS    HTC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1359)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Marcello Bento Soares, Ph.D.
            cDNA Library Preparation: Soares Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 66 Row: c Column: 13
            This clone has the following problem: no polyA-tail.
FEATURES             Location/Qualifiers
     source          1. .1359
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1179696"
                     /tissue_type="Mammary gland"
```

```
                          /clone_lib="Soares_mammary_gland_NbMMG"
                          /lab_host="DH10B"
                          /note="Vector: pT7T3D-Pac"
BASE COUNT       505 a    296 c    304 g    254 t
ORIGIN

Alignment Scores:
Pred. No.:               2.07e-153      Length:          1359
Score:                   2170.50        Matches:         419
Percent Similarity:      95.14%         Conservative:    12
Best Local Similarity:   92.49%         Mismatches:      19
Query Match:             30.18%         Indels:          3
DB:                      11             Gaps:            2

US-09-512-581B-2 (1-1391) x BC032988 (1-1359)

Qy        939 TyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAla 958
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 TACATGGCCATCTGTGCTCTTTGTGCCAAAGACCCTGTGAAAGAGAGGCGAGCCCATGCT 62

Qy        959 ArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAla 978
              |||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db         63 AGACAGTGCTTGGTGAAGAACATCACTGTGAGGAGGGAGTACCTGAAGCAGCATGCAGCT 122

Qy        979 ValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeu 998
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 GTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTT 182

Qy        999 LeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLys 1018
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 TTGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTGAAA 242

Qy       1019 GluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAla 1038
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 GAATGCCTTTGGTTTGTTCTGGAGATATTGATGGCTAAAAATGAAAACAACAGCCATGCA 302

Qy       1039 PheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAsp 1058
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 TTTATCAGAAAAATGGTAGAAAATATTAAACAGACAAAAGATGCTCAAGGACCAGATGAT 362

Qy       1059 AlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSer 1078
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 ACAAAAATGAATGAAAAATTGTACACCGTGTGTGATGTTGCCATGAACATCATCATGTCC 422

Qy       1079 LysSerThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePhe 1098
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 AAGAGCACCACGTACAGCCTGGAGTCTCCTAAGGACCCCGTGCTGCCAGCTCGGTTTTTC 482

Qy       1099 ThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSer 1118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 ACCCAGCCTGACAAGAATTTTAGTAACACCAAAAATTACCTGCCTCCAGAAATGAAATCA 542

Qy       1119 PhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeu 1138
              |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        543 TTTTTCACTCCTGGAAAACCTAAAACAGCCAATGTTCTCGGAGCAGTTAATAAGCCACTT 602

Qy       1139 SerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAla 1158
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 TCATCAGCAGGCAAACAGTCTCAGACCAAATCATCAAGAATGGAAACTGTGAGCAACGCA 662

Qy       1159 SerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGlu 1178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 AGCAGCAGCTCCAACCCAAGCTCTCCTGGAAGGATCAAGGGGAGGCTTGATAGCTCTGAA 722

Qy       1179 MetAspHisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSer 1198
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        723 ATGGATCACAGTGAAAATGAAGATTATACAATGTCTTCACCTTTGCCAGGAAAAAAAAGT 782

Qy       1199 AspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLys 1218
              ||||||||:::|||   ||||||          ||||||||||||||||||    ||||||
Db        783 GACAAGAGAGAAGACCCTGATCTT------TCTGAGTTGGAGAAGCCTAGAAGTCGGAAA 836

Qy       1219 LysThrProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuVal 1238
              |||   |||||||||:::   ||||||||||||||||||||||||||||||||||||||
Db        837 AAAGCACCTGTCACAGACCCTGAAGAGAAATTAGGTATGGATGACCTAACTAAGTTGGTA 896

Qy       1239 GlnGluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGlu 1258
              |||||||||||||||||||||||||||||||   |||||||||||||    |||||||||:::
Db        897 CAGGAACAGAAACCTAAAGGCAGTCAGCGAGGCCGGAAAAGAGGCCGTACAGCTTCAGAC 956

Qy       1259 SerAspGluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGlu 1278
              ||||||||||||||||||||||||||||||||||||   |||||:::::||||||||||||
Db        957 TCAGACGAGCAGCAGTGGCCTGAGGAGAAGAGGCACAAAGAGGAGCTCCTGGAAAATGAG 1016

Qy       1279 AspGluGlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGly 1298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1017 GATGAGCAGAACAGCCCACCAAAAAAGGGCAAAAGAGGCAGGCCACCAAAACCTCTTGGT 1076

Qy       1299 GlyGlyThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLys 1318
              |||||||||   ||||||||||||||||||||||||||||||||||||||:::|||||||||
Db       1077 GGGGGACATCGAAGGAAGAGCCCACAATGAAAACATCCAAGAAAGGAAACAAGAAAAAA 1136

Qy       1319 SerGlyProProAlaProGluGluGluGluGluGluArgGlnSerGlyAsnThrGlu 1338
                     ||||||      :::::::::||||||||||||||||    ||||||||||||
Db       1137 CTTGTACCTCCTGTAGTAGACGACGATGAAGAAGAAGAAAGACAAATTGGAAACACAGAA 1196

Qy       1339 GlnLysSerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSer 1358
                 ||||||||||||||||||||||   |||:::||||||||||||||||||||||||||
Db       1197 CATAAGTCAAAAAGCAAACAGCACCGCACATCAAAGAGAGCACAACAGAGAGCAGAATCT 1256

Qy       1359 ProGluSerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgPro 1378
              ||||||:::||||||:::||||||||||||||||||||||||||||||||||||||||||
Db       1257 CCTGAAACAAGTGCAGTTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCA 1316

Qy       1379 SerLysThrProSerProSerGln---ProLysLysAsn 1390
              ||||||    ||||||||||||||||    ||||||||||||
Db       1317 TCAAAAGCACCATCACCATCACAACCCCCCAAAAAAAAC 1355

RESULT 6
CA980171
LOCUS       CA980171                 926 bp    mRNA    linear   EST 06-JAN-2003
DEFINITION  AGENCOURT_11295822 NIH_MGC_164 Mus musculus cDNA clone
            IMAGE:30145391 5', mRNA sequence.
ACCESSION   CA980171
VERSION     CA980171.1  GI:27512825
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 926)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. David Rowe and Dr. Mina
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM0059  row: f  column: 24
            High quality sequence stop: 620.
FEATURES             Location/Qualifiers
     source          1..926
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30145391"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_164"
                     /note="Vector: pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2:
                     NotI; Non-normalized full-length enriched library from
                     pooled mouse embryonic limb, maxilla and mandible, day
```

```
              10.5 and 11.5 (size selected for the 0.5-1 kb fragments)
              Cloned directionally, priming method: Oligo-dT.  cDNA
              enrichment: >1k bp. Average insert size 1.8k bp. Priming
              sequence: 5'GACTAGTTCTAGATCGCGAGCGGCCGCCC(T) 3'. Tissue
              contributed by, David Rowe. Library constructed by ResGen,
              Invitrogen Corp."
BASE COUNT      332 a    199 c    222 g    172 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              3.98e-95        Length:         926
Score:                  1395.00         Matches:        277
Percent Similarity:     94.02%          Conservative:   6
Best Local Similarity:  92.03%          Mismatches:     16
Query Match:            19.39%          Indels:         3
DB:                     14              Gaps:           0

US-09-512-581B-2 (1-1391) x CA980171 (1-926)

Qy       1015 LysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsn 1034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         26 AAAGATGTGAAAGAATGCCTTTGGTTTGTTCTGGAGATATTGATGGCTAAAAATGAAAAC 85

Qy       1035 AsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGln 1054
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 AACAGCCATGCATTTATCAGAAAAATGGTAGAAAATATTAAACAGACAAAAGATGCTCAA 145

Qy       1055 GlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsn 1074
              |||||||||||  |||||||||||||||||||||||||||||||||||||||||||||||
Db        146 GGACCAGATGATACAAAAATGAATGAAAAATTGTACACCGTGTGTGATGTTGCCATGAAC 205

Qy       1075 IleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProValLeuPro 1094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 ATCATCATGTCCAAGAGCACCACGTACAGCCTGGAGTCTCCTAAGGACCCCGTGCTGCCA 265

Qy       1095 AlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProPro 1114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 GCTCGGTTTTTCACCCAGCCTGACAAGAATTTTAGTAACACCAAAAATTACCTGCCTCCA 325

Qy       1115 GluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaVal 1134
              ||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db        326 GAAATGAAATCATTTTTCACTCCTGGAAAACCTAAAACAGCCAATGTTCTCGGAGCAGTT 385

Qy       1135 AsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThr 1154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 AATAAGCCACTTTCATCAGCAGGCAAACAGTCTCAGACCAAATCATCAAGAATGGAAACT 445

Qy       1155 ValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeu 1174
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        446 GTGAGCAACGCAAGCAGCAGCTCCAACCCAAGCTCTCCTGGAAGGATCAAGGGGAGGCTT 505

Qy       1175 AspSerSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSerProLeuPro 1194
              ||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db        506 GATAGCTCTGAAATGGATCACAGTGAAAATGAAGATTATAC-ATGTCTTCACCTTTGCCA 564

Qy       1195 GlyLysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysPro 1214
              ||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||
Db        565 GGAAAAAAAAGTGACAAGAGAGAAGACCCTGATCTTGTAAGGTCTGAGCTGGAGAAGCCT 624

Qy       1215 ArgGlyArgLysLysThrProValThrGluGlnGluGluLysLeuGlyMetAspAspLeu 1234
              |||  ||||||||  ||||||||:::  ||||||||||||||||||||||||||||||||
Db        625 AGAAGTCGGAAAAAAGCACCTGTCACAGACCCTGAAGAGAAATTAGGTATGGATGACCTA 684

Qy       1235 ThrLysLeuValGlnGluGlnLysProLysGlySerGlnArgSerArg-LysArgGlyHi 1254
              |||||||||||||||||||||||||||||||||||||||||    ||| |||||||||
Db        685 ACTAAGTTGGTACAGGAACAGAAACCTAAAGGCAGTCAGCGAGGCCGGAAAAAGAGGCCG 744

Qy       1254 sThrAlaSerGluSerAspGluGlnGlnTrpProGluGluLysArgLeuLysGluAspIl 1274
              ||||||||:::||||||||||||||||||||||||||||||||    ||||||::::
Db        745 TACAGCTTCAGACTCAGACGAGCAGCAGTGGCCTGAGGAGAAGAGGCACAAAGAGGAGCT 804

Qy       1274 eLeuGluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgProPr 1294
              :|||||||||||||||||||    .    |||||||||||||||||||||||||||||||
Db        805 CCTGGAAAATGAGGATGAGCAGAAACAGCCCCCAAAAAAGGGCAAAAGAGGCAGGCCCAC 864

Qy       1294 oLysProLeuGly-GlyGlyThrProLysGluGluProThrMetLysThrSerLysLys 1313
              ||||||||||| ||||||||   |||   ||||||||   :::||||||||||||
Db        865 AAAACCTCTTGGNTGGGGGGACATCGAAGGGAGAACCCACCAATGAAACTTCCAAGAAA 923

RESULT 7
CA324097
LOCUS       CA324097                 786 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  UI-M-FY0-cco-m-16-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
            IMAGE: 6822425 5', mRNA sequence.
ACCESSION   CA324097
VERSION     CA324097.1  GI:24542195
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 786)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..786
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6822425"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FY0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      256 a    171 c    165 g    191 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:              6.18e-90        Length:         786
Score:                  1325.00         Matches:        256
Percent Similarity:     97.71%          Conservative:   0
Best Local Similarity:  97.71%          Mismatches:     6
Query Match:            18.42%          Indels:         0
DB:                     14              Gaps:           0

US-09-512-581B-2 (1-1391) x CA324097 (1-786)
```

```
Qy        932 ArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProVal 951
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CGCTTACGGCTTCCCCTTGAGTACATGGCCATCTGTGCTCTTTGTGCCAAAGACCCTGTG 60

Qy        952 LysGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGlu 971
              |||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db         61 AAAGAGAGGCGAGCCCATGCTAGACAGTGCTTGGTGAAGAACATCACTGTGAGGAGGGAG 120

Qy        972 TyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrVal 991
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TACCTGAAGCAGCATGCAGCTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATGTT 180

Qy        992 ValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIle 1011
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTTCCATATACAATTCACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATATT 240

Qy       1012 GluGlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLys 1031
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAACAACTTAAAGATGTGAAAGAATGCCTTTGGTTTGTTCTGGAGATATTGATGGCTAAA 300

Qy       1032 AsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLys 1051
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AATGAAAACAACAGCCATGCATTTATCAGAAAAATGGTAGAAAATATTAAACAGACAAAA 360

Qy       1052 AspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspVal 1071
              |||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        361 GATGCTCAAGGACCAGATGATACAAAAATGAATGAAAAATTGTACACCGTGTGTGATGTT 420

Qy       1072 AlaMetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspPro 1091
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCCATGAACATCATCATGTCCAAGAGCACCACGTACAGCCTGGAGTCTCCTAAGGACCCC 480

Qy       1092 ValLeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyr 1111
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTGCTGCCAGCTCGGTTTTTCACCCAGCCTGACAAGAATTTTAGTAACACCAAAAATTAC 540

Qy       1112 LeuProProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeu 1131
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        541 CTGCCTCCAGAAATGAAATCATTTTTCACTCCTGGAAAACCTAAAACAGCCAATGTTCTC 600

Qy       1132 GlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArg 1151
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        601 GGAGCAGTTAATAAGCCACTTTCATCAGCAGGCAAACAGTCTCAGACCANATCATCAAGA 660

Qy       1152 MetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLys 1171
              |||||||||||||||||||||||||||||||||   |||   ||||||||||||||||||
Db        661 ATGGAAACTGTGAGCAACGCAAGCAGCAGCTCCANCCCANGCTCTCCTGGAAGGATCAAG 720

Qy       1172 GlyArgLeuAspSerSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSer 1191
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGGAGGCTTGATAGCTCTGAAATGGATCACAGTGAAAATGAAGATTATACAATGTCTTCA 780

Qy       1192 ProLeu 1193
              ||||||
Db        781 CCTTTG 786
```

```
RESULT 8
CA324284
LOCUS       CA324284                 779 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  UI-M-FY0-ccp-a-07-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
            IMAGE: 6822512 5', mRNA sequence.
ACCESSION   CA324284
VERSION     CA324284.1  GI:24542382
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 779)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .779
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6822512"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FY0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      241 a    166 c    174 g    197 t     1 others
ORIGIN

Alignment Scores:
Pred. No.:              6.95e-89       Length:         779
Score:                  1311.00        Matches:        255
Percent Similarity:     98.46%         Conservative:   0
Best Local Similarity:  98.46%         Mismatches:     4
Query Match:            18.23%         Indels:         1
DB:                     14             Gaps:           0

US-09-512-581B-2 (1-1391) x CA324284 (1-779)

Qy        863 AspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAla 882
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 GATTTGACAGAACAAGGGAAAATTAGTAAACCAGATATGTCACGCCTGAGACTTGCTGCT 62

Qy        883 GlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGlu 902
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 GGGAGTGCTATTGTGAAGCTGGCACAGGAGCCCTGTTACCACGAGATCATTACACTGGAG 122

Qy        903 GlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPhe 922
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 CAGTACCAGCTGTGTGCATTAGCCATCAATGATGAGTGTTATCAAGTCAGGCAGGTGTTC 182

Qy        923 AlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIle 942
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 GCTCAGAAACTTCACAAAGGCCTTTCCCGCTTACGGCTTCCCCTTGAGTACATGGCCATC 242

Qy        943 CysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeu 962
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 TGTGCTCTTTGTGCCAAAGACCCTGTGAAAGAGAGGCGAGCCCATGCTAGACAGTGCTTG 302

Qy        963 ValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLys 982
              ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db        303 GTGAAGAACATCACTGTGAGGAGGGAGTACCTGAAGCAGCATGCAGCTGTTAGTGAAAAA 362
```

```
Qy        983 LeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAsp 1002
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 TTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCACATGAC 422

Qy       1003 ProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrp 1022
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 CCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTGAAAGAATGCCTTTGG 482

Qy       1023 PheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLys 1042
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 TTTGTTCTGGAGATATTGATGGCTAAAAATGAAAACAACAGCCATGCATTTATCAGAAAA 542

Qy       1043 MetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsn 1062
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        543 ATGGTAGAAAATATTAAACAGACAAAAGATGCTCAAGGACCAGATGATACAAAAATGAAT 602

Qy       1063 GluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSerThrThr 1082
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 GAAAAATTGTACACCGTGTGTGATGTTGCCATGAACATCATCATGTCCAAGAGCACCACG 662

Qy       1083 TyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGlnProAsp 1102
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 TACAGCCTGGAGTCTCCTAAGGACCCCGTGCTGCCAGCTCGGTTTTTCACCCAGCCTGAC 722

Qy       1103 LysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePheThr 1121
              |||||||||||||||||   ||||||||||||||||||||   ||||||||||||
Db        723 AAGAATTTTAGTAACACCANAAATTACCTGCCTCCAGAAATGAA-TCATTTTTCACT 778
```

RESULT 9
BU703488

```
LOCUS       BU703488                 786 bp    mRNA    linear   EST 09-OCT-2002
DEFINITION  UI-M-FO0-bzo-g-20-0-UI.r1 NIH_BMAP_FO0 Mus musculus cDNA clone
            IMAGE: 6405163 5', mRNA sequence.
ACCESSION   BU703488
VERSION     BU703488.1  GI:23629371
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 786)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..786
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6405163"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 12.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FO0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is TGAGAGAGCC. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      248 a    143 c    152 g    242 t      1 others
ORIGIN
```

```
Alignment Scores:
Pred. No.:                6.16e-88       Length:          786
Score:                    1298.50        Matches:         254
Percent Similarity:       98.08%         Conservative:    1
Best Local Similarity:    97.69%         Mismatches:      4
Query Match:              18.05%         Indels:          1
DB:                       13             Gaps:            1
```

US-09-512-581B-2 (1-1391) x BU703488 (1-786)

```
Qy          1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProProGlyValLys 20
              |||||||||    :::   |||||||||||||||||||||||||||||||||||||||||
Db         10 ATGGCTCATCAAGAA---AGGACCAACGATGGGAAAATTACTTACCCTCCTGGAGTCAAG 66

Qy         21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValValLys 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 GAAATCTCAGATAAAATCTCTAAAGAGGAGATGGTGAGGCGGTTAAAGATGGTTGTAAAA 126

Qy         41 ThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 ACTTTCATGGACATGGACCAGGACTCTGAAGAGGAAAAGGAACTTTATCTAAACCTAGCT 186

Qy         61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
              ||||||||||||||||||||||||||||||||||   |||||||||||||||||||||||
Db        187 TTACATCTTGCTTCTGACTTCTTCCTCAAGCATCCTGATAAAGATGTTCGTTTACTGGTG 246

Qy         81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 GCTTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAGGCTCCTTACACGTCTCCC 306

Qy        101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 GATAAACTAAAGGATATATTTATGTTTATAACAAGGCAACTAAAGGGACTAGAAGATACA 366

Qy        121 LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 AAGAGCCCTCAATTTAATAGGTATTTTTATTTACTTGAGAACATTGCATGGGTTAAATCA 426

Qy        141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATCTTTACTCAACTTTACAGAACA 486

Qy        161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        487 TTATTCTCAGTTATAAACAATGGCCACAATCAGAAAGTTCATATGCACATGGTGGACCTC 546

Qy        181 MetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        547 ATGAGCTCTATCATTTGTGAAGGTGATACCGTATCTCAGGAGCTCTTAGATACAGTTTTA 606

Qy        201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        607 GTAAATCTGGTACCTGCCCATAAGAACTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 666

Qy        221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
              ||||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db        667 TTGCTGAAGAGGACTGCTCAAGCTATTGAACCATATATTACCAATTTTTTAACCAGGTT 726

Qy        241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeu 260
              |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        727 CTGATGCTTGNGAAAACATCTATCAGTGATTTGTCTGAGCATGTCTTTGATTTGATCTTG 786

RESULT 10
BU363024
LOCUS       BU363024                782 bp    mRNA    linear   EST 28-NOV-2002
DEFINITION  603790485F1 CSEQCHN72 Gallus gallus cDNA clone ChEST753i21 5', mRNA
            sequence.
ACCESSION   BU363024
VERSION     BU363024.1  GI:25871025
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 782)
  AUTHORS   Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
  TITLE     A Comprehensive Collection of Chicken cDNAs
  JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
  MEDLINE   22335534
   PUBMED   12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology (UMIST
            )
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
FEATURES             Location/Qualifiers
     source          1..782
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="Compton Line 15I"
                     /db_xref="taxon:9031"
                     /clone="ChEST753i21"
                     /sex="Female"
                     /tissue_type="cerebrum"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN72"
                     /note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site_2: NotI; This normalized library was
                     constructed from 1 million independent clones. cDNA
                     synthesis was initiated using an oligo(dT) primer, using
                     methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with EcoRI
                     , size-selected, and cloned into the NotI and EcoRI
                     compatible sites of a custom modified MCS of the
                     pBluescript (KS+) vector. The library was normalized in 2
                     rounds using conditions adapted from Soares et al., PNAS
                     (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
                     (1996): 791, except that a significantly longer
                     reannealing hybridization was used."
BASE COUNT      279 a    146 c    160 g    197 t
ORIGIN

Alignment Scores:
Pred. No.:              1.12e-87      Length:         782
Score:                  1295.00       Matches:        253
Percent Similarity:     98.85%        Conservative:   5
Best Local Similarity:  96.93%        Mismatches:     2
Query Match:            18.00%        Indels:         2
DB:                     13            Gaps:           0

US-09-512-581B-2 (1-1391) x BU363024 (1-782)

Qy        856 ThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMet 875
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 ACAATATTACACAGTGATGGAGACTTGACAGAGCAGGGGAAAATTAGTAAACCTGATATG 62

Qy        876 SerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyr 895
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 TCTCGATTGAGGTTGGCCGCAGGTAGTGCTATTGTGAAGCTGGCACAAGAACCTTGTTAC 122

Qy        896 HisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCys 915
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 CATGAAATCATCACCTTAGAACAGTACCAGCTGTGTGCACTAGCCATAAACGATGAATGT 182

Qy        916 TyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeu 935
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        183 TACCAAGTGAGGCAAATATTTGCTCAGAAACTTCACAAAGGGCTTTCTAGACTACGGCTG 242

Qy        936 ProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArg 955
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 CCACTAGAATATATGGCTATTTGTGCACTGTGTGCAAAAGATCCAGTGAAAGAGAGAAGA 302

Qy        956 AlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGln 975
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 GCTCATGCTAGGCAGTGCCTTGTGAAGAACATAAATGTCAGAAGGGAATATCTGAAGCAA 362

Qy        976 HisAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThr 995
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 CATGCAGCAGTTAGTGAGAAACTGTTGTCGCTTCTACCAGAGTATGTTGTTCCTTATACA 422

Qy        996 IleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeuLys 1015
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 ATACATCTTCTAGCTCATGACCCAGATTATGTCAAAGTCCAGGACATCGAACAACTAAAG 482

Qy       1016 AspValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsn 1035
              |||:::||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        483 GACATTAAAGAGTGTTTGTGGTTCATACTGGAAATATTGATGGCTAAAAATGAGAACAAC 542

Qy       1036 SerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGly 1055
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 AGTCATGCATTTATCAGAAAAATGGTAGAAAATATTAAACAGACTAAAGATGCTCAAGGA 602

Qy       1056 ProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIle 1075
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 CCAGATGATGCAAAAATGAATGAAAAACTCTACACAGTCTGTGATGTAGCAATGAATATC 662

Qy       1076 IleMetSerLysSerThrThr-TyrSerLeuGluSerProLysAspProValLeuProAl 1095
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        663 ATTATGTCAAAGAGTACAACCATACAGTTTGGAATCCCCTAAAGATCCTGTACTTCCAGC 722

Qy       1095 aArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGl 1115
              ||||:::  |||||||||||||||   |||||||||||||||||||||||||||||||::
Db        723 ACGATATGTTACTCAACCTGACAAGAA-TTTAGTAACACCAAAAATTATCTTCCTCCAAA 781

Qy       1115 u 1115
              :
Db        782 A 782

RESULT 11
BU914579
LOCUS       BU914579                943 bp    mRNA    linear   EST 17-OCT-2002
DEFINITION  AGENCOURT_10496299 NICHD_XGC_OO1 Xenopus laevis cDNA clone
            IMAGE:6641106 5', mRNA sequence.
ACCESSION   BU914579
VERSION     BU914579.1  GI:24096493
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 943)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.
```

```
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14224  row: f  column: 18
            High quality sequence stop: 746.
FEATURES             Location/Qualifiers
     source          1..943
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:6641106"
                     /tissue_type="oocytes"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NICHD_XGC_OO1"
                     /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 2.2 kb. Constructed by Life Technologies."
BASE COUNT      297 a    182 c    215 g    249 t
ORIGIN

Alignment Scores:
Pred. No.:              7.38e-87       Length:        943
Score:                  1285.50        Matches:       261
Percent Similarity:     87.38%         Conservative:  16
Best Local Similarity:  82.33%         Mismatches:    30
Query Match:            17.87%         Indels:        10
DB:                     13             Gaps:          5

US-09-512-581B-2 (1-1391) x BU914579 (1-943)

Qy        744 PheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeuGluHisLeu 763
              ||||||||||||||||||||||||||||||||||||||||||||    |||    |||  |||
Db          1 TTTGCACAGATATTTGAGCCCTTACATAAAAGCTTGGACCCTGGCAACCCTGAGCAACTC 60

Qy        764 IleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAla 783
              ||||||   ||||||:::||||||||||||||.|||   |||||||||||||||||   |||
Db         61 ATCACATCACTCGTCTCCATTGGCCACATAGCACAGCTTGCTCCAGATCAGTTCACTGCT 120

Qy        784 ProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeu 803
              |||   ||||||   ||||||||||||:::||||||||||||||||||     ||||||||
Db        121 CCTTTAAAATCCATGGTTGCCACCTTTGTGGTTAAAGATTTGTTGATGACAGATCGGCTT 180

Qy        804 ProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerProGluThrMet 823
              ||||||||||||||||||||||||||||||     |||:::|||||||||    ||||||
Db        181 CCAGGTAAAAAGACAACAAAACTCTGGGTGTCAGATGATGAGGTGTCGACAGAAACTAAG 240

Qy        824 ValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnHis 843
              ||||||||||||||||||||||||||||||||||||||||||:::||||||||||||||||
Db        241 GTTAAAATTCAAGCAATTAAAATGATGGTGCGGTGGCTGGTGGGAATGAAAAATAATCTC 300

Qy        844 SerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAsp 863
              ||||||||||||   ||||||||||||||||     |||||||||||:::||||||||||
Db        301 AGTAAATCCGGGAATTCTACCTTACGACTCCTAATGGCAATATTGCACACTGACGGAGAC 360

Qy        864 LeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGly 883
              |||||||||   ||||||:::||||||||||||||||||||||||||||||||||||||||
Db        361 TTGACAGAACATGGAAAGCTCAGTAAACCAGACATGTCACGGCTGAGACTGGCTGCTGGC 420

Qy        884 SerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGluGln 903
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGTGCCATTGTTAAACTAGCGCAGGAGCCATGTTACCATGAAATTATTACCTTAGAACAG 480

Qy        904 TyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAla 923
              ||||||||||||||||||   ||||||||||||||||||||||||||||||:::||||||
Db        481 TATCAGCTGTGTGCATTGGTCATCAATGATGAGTGTTACCAGGTTAGACAGTTGTTTGCT 540

Qy        924 GlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCys 943
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAAAAATTCACAAAGGGCTCTCCAGATTAAGACTTCCATTAGAATATATGGCCATCTGT 600

Qy        944 AlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuVal 963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCTCTGTGTGCAAAAGATCCTGTGAAAGAGAGACGCGCTCACGCACGGCAGTGTCTGGTG 660

Qy        964 LysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeu 983
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AAAAATATTAATGTCAGAAGAGAGTATCTGAAACAGCATGCAGCTGTGAGTGAAAAATTG 720

Qy        984 LeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspPro 1003
                 |||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        721 TTCTCTCTTCTGCCTGAATATGTGGTTCCATATACTGTTCATTTGCTTGCACATGACCCA 780

Qy       1004 AspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrp--- 1022
              |||      :::|||||||||||||||||||||||||||:::||||||         |||
Db        781 GATTTATGTGAAGTACAAGATATAGAGCAGCTAAAGGACATCAAGGAG------TGGCCT 834

Qy       1023 ------PheValLeu---GluIleLeuMetAlaLysAsnGluAsnAsn-SerHisAlaPh 1039
                    |||||||||   :::   |||      :::|||||||||||| ||||||   ||
Db        835 GTGGGGTTCGTGCTGGGAGATTCTTTATGTCAAAGAAATGAAAATAACAAGCCATGGCTT 894

Qy       1039 e---IleArgLysMet------ValGluAsnIleLysGlnThrLysAsp 1052
              |    ||||||               |||:::    :::|||||||:::
Db        895 TTAATTTCGGAAAAATGGGTAGAAATATAATTTAGGCAAAACAAAAGAA 943

RESULT 12
CA987660
LOCUS       CA987660                 857 bp    mRNA    linear   EST 06-JAN-2003
DEFINITION  AGENCOURT_11108371 NICHD_XGC_Embl Xenopus laevis cDNA clone
            IMAGE:6864353 5', mRNA sequence.
ACCESSION   CA987660
VERSION     CA987660.1  GI:27520335
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 857)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14484  row: p  column: 16
            High quality sequence stop: 670.
FEATURES             Location/Qualifiers
     source          1..857
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:6864353"
                     /tissue_type="embryo (stage 10)"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NICHD_XGC_Embl"
                     /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 1.55 kb. Constructed by Life Technologies. Note: This
                     is a Xenopus Gene Collection (XGC) library."
BASE COUNT      275 a    160 c    192 g    230 t
ORIGIN

Alignment Scores:
Pred. No.:              1.06e-85       Length:        857
Score:                  1269.50        Matches:       249
Percent Similarity:     91.81%         Conservative:  9
```

```
Best Local Similarity:  88.61%        Mismatches:    21
Query Match:            17.65%        Indels:         2
DB:                     14            Gaps:           1

US-09-512-581B-2 (1-1391) x CA987660 (1-857)

Qy        784 ProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeu 803
              |||  ||||||  |||||||||||:::||||||||||||||||   |||||||||
Db          3 CCTTTAAAATCCATGGTTGCCACCTTTGTGGTTAAAGATTTGTTGATGACAGATCGGCTT 62

Qy        804 ProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerProGluThrMet 823
              ||||||||||||||||||||||||||||||  |||:::||||||||||   ||||||
Db         63 CCAGGTAAAAAGACAACAAAACTCTGGGTGTCAGATGATGAGGTGTCGACAGAAACTAAG 122

Qy        824 ValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnHis 843
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 GTTAAAATTCAAGCAATTAAAATGATGGTGCGGTGGCTGCTGGGAATGAAAAATAATCTC 182

Qy        844 SerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAsp 863
              |||||||||||  ||||||||||||||||       |||||||||:::|||||||||
Db        183 AGTAAATCCGGGAATTCTACTTTACGACTCCTAATGGCAATATTGCACACTGACGGAGAC 242

Qy        864 LeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGly 883
              |||||||||  ||||||:::||||||||||||||||||||||||||||||||||||||
Db        243 TTGACAGAACATGGAAAGCTCAGTAAACCAGACATGTCACGGCTGAGACTGGCTGCTGCC 302

Qy        884 SerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGluGln 903
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 AGTGCCATTGTTAAACTAGCGCAGGAGCCATGTTACCATGAAATTATTACCTTAGAACAG 362

Qy        904 TyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAla 923
              ||||||||||||||||||   |||||||||||||||||||||||||||||:::||||||
Db        363 TATCAGCTATGTGCATTGGTCATCAATGATGAGTGTTACCAGGTTAGACAGTTGTTTGCT 422

Qy        924 GlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCys 943
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 CAAAAAATTCACAAAGGGCTCTCCAGATTAAGACTTCCATTAGAATATATGGCCATCTGT 482

Qy        944 AlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuVal 963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 GCTCTATGTGCAAAAGATCCTGTGAAAGAGAGGCGCGCTCACGCACGGCAGTGTCTGGTG 542

Qy        964 LysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeu 983
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 AAAAATATTAATGTCAGAAGAGAGTATCTGAAACAGCATGCAGCTGTGAGTGAAAAATTG 602

Qy        984 LeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspPro 1003
                 ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        603 TTCTCTCTTCTGCCTGAATATGTGGTTCCATATACTGTTCATTTGCTTGCACATGACCCA 662

Qy       1004 AspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys-LeuTrpPh 1023
              ||||||||||||||||||||||||||||||||||||||:::|||||||||  |||||||
Db        663 GATTATGTGAAAGTACAAGATATAGAGCAGCTAAAGGACATCAAAGAGTGCCCTGTGGTT 722

Qy       1023 eValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMe 1043
              ||||||||||||||||||:::|||||||||||||||||||||||||||||     |||||
Db        723 CGTGCTGGAAATTCTTATGTCAAAGAATGAAAATAACAGCCATGCTTTTATTCCGAAAAT 782

Qy       1043 tValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGl 1063
              |||||||  ||||||  |||||||||              |||||||||  |||  |||||
Db        783 GGTAGAATATATTAAGCCAACCAAAGATGGCCCAAATCCTGATGACCAGAAG---AATGA 839

Qy       1063 u 1063
              |
Db        840 A 840
```

RESULT 13
BU916668

```
LOCUS       BU916668                882 bp    mRNA    linear   EST 17-OCT-2002
DEFINITION  AGENCOURT_10496784 NICHD_XGC_OO1 Xenopus laevis cDNA clone
            IMAGE:6643610 5', mRNA sequence.
ACCESSION   BU916668
VERSION     BU916668.1  GI:24098582
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 882)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14230  row: o  column: 02
            High quality sequence stop: 683.
FEATURES             Location/Qualifiers
     source          1..882
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:6643610"
                     /tissue_type="oocytes"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NICHD_XGC_OO1"
                     /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 2.2 kb. Constructed by Life Technologies."
BASE COUNT      310 a    169 c    192 g    210 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              3.69e-85      Length:         882
Score:                  1262.50       Matches:        246
Percent Similarity:     91.35%        Conservative:   18
Best Local Similarity:  85.12%        Mismatches:     25
Query Match:            17.55%        Indels:         1
DB:                     13            Gaps:           0

US-09-512-581B-2 (1-1391) x BU916668 (1-882)

Qy        956 AlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGln 975
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 GCTCACGCACGGCAGTGTCTGGTGAAAAATATTAATGTCAGAAGAGAGTATCTGAAACAG 62

Qy        976 HisAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThr 995
              ||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db         63 CATGCAGCTGTGAGTGAAAAATTGTTCTCTCTTCTGCCTGAATATGTGGTTCCATATACT 122

Qy        996 IleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeuLys 1015
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 GTTCATTTGCTTGCACATGACCCAGATTATGTGAAAGTACAAGATATAGAGCAGCTAAAG 182

Qy       1016 AspValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsn 1035
              |||:::||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        183 GACATCAAAGAGTGCCTGTGGTTCGTGCTGGAGATTCTTATGTCAAAGAATGAAAATAAC 242

Qy       1036 SerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGly 1055
              |||||||||||||||||||||||||||  |||||||||||||||||   |||
Db        243 AGCCATGCTTTTATTCGGAAAATGGTAGAATATATTAAGCAAACAAAAGATGGCCAAAAC 302

Qy       1056 ProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIle 1075
              |||||||||   |||||||||||||||:::||||||||||||||||||||||||||||||
Db        303 CCCGATGACCAGAAGATGAATGAAAAAATGTACACAGTGTGTGATGTGGCAATGAATATA 362

Qy       1076 IleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProValLeuProAla 1095
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      363 ATTATTTCAAAGAGCACAACATACAGCTTAGAGTCCCCGAAGGATCCTGTATTGCCTGCA 422

Qy     1096 ArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGlu 1115
            |||||||||||||||||||||||||||||||||||||||||||||||||||||    |||
Db      423 CGTTTCTTCACTCAGCCAGATAAGAATTTCAGCAACACAAAAAATTATCTTCCCGCTGAA 482

Qy     1116 MetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsn 1135
            :::||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db      483 TTGAAGTCCTTCTTTACACCAGGAAAGCCAAAATCTACTAATGTTCTGGGTGCAGTCAAT 542

Qy     1136 LysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrVal 1155
            |||||||||||||||||||||||||        :::||||||||||||||||||||||||
Db      543 AAACCCTTATCGAGCGCAGGAAAGCAAATGCTATCCAAGTCATCTAGGATGGAGACTGTA 602

Qy     1156 SerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAsp 1175
            ||||||||||||||   |||||||||||||||||||||:::|||||||||||||||||||
Db      603 AGCAATGCAAGCAGCGGTTCTAACCCAAGCTCTCCAGGCAAGATAAAAGGAAGACTTGAC 662

Qy     1176 SerSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSerProLeuProGly 1195
            |||   |||:::|||   |||||||||||||||||||||||||||||||||    |||
Db      663 AGTATGGAATTGGATCAAAGTGAAAATGAAGACTATACGATGTCTTCTCCTCTTTCTGGG 722

Qy     1196 LysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArg 1215
            |||||||||||||||||||||||||||||||||:::::::||||||||||||||||||||
Db      723 AAAAAAAGTGACAAAAGGGACGACTCGGATCTCTTGAAGTCTGAGTTAGAAAAGCCAAGG 782

Qy     1216 GlyArgLysLysThrProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThr 1235
               |||:::|||         :::|||::::::   |||   ||||||::::::
Db      783 CG-AGGGAAAAACAATCTCCTATAGACCAAGATGACAGTTTAAGCATGGATGAAATTATC 841

Qy     1236 LysLeuValGlnGluGlnLysProLys 1244
            |||      |||      |||   |||
Db      842 AAACCAGCACAGGGAACCAAAATCAAG 868

RESULT 14
BU914025
LOCUS       BU914025                 931 bp    mRNA    linear   EST 17-OCT-2002
DEFINITION  AGENCOURT_10494702 NICHD_XGC_OO1 Xenopus laevis cDNA clone
            IMAGE:6640480 5', mRNA sequence.
ACCESSION   BU914025
VERSION     BU914025.1  GI:24095939
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 931)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14222  row: 1  column: 16
            High quality sequence stop: 552.
FEATURES             Location/Qualifiers
     source          1..931
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:6640480"
                     /tissue_type="oocytes"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NICHD_XGC_OO1"
                     /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 2.2 kb. Constructed by Life Technologies."
BASE COUNT      286 a    173 c    202 g    269 t     1 others
ORIGIN

Alignment Scores:
Pred. No.:              2.33e-83      Length:          931
Score:                  1239.00       Matches:         253
Percent Similarity:     93.88%        Conservative:    8
Best Local Similarity:  91.01%        Mismatches:      14
Query Match:            17.23%        Indels:          4
DB:                     13            Gaps:            0

US-09-512-581B-2 (1-1391) x BU914025 (1-931)

Qy        1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProProGlyValLys 20
            ||||||   ||||||    :::   ||||||||||||||||||||||||||||||||||||
Db       98 ATGGCTCTCTCCAAGGCCAAGGGCAACGATGGAAAAATCACGTATCCCCCAGGTGTCAAG 157

Qy       21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValValLys 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      158 GAAATTTCAGACAAGATTTCCAAAGAAGAAATGGTTAGACGGTTAAAGATGGTTGTGAAA 217

Qy       41 ThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAla 60
            ||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db      218 ACCTTCATGGATATGGATCAAGACTCTGAAGAAGAAAAGGAGCAGTATTTGAACCTTGCT 277

Qy       61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
            ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||
Db      278 TTACACTTAGCCTCTGACTTTTTCCTCAAACATCCTGACAAAGATGTACGATTGCTTGTA 337

Qy       81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      338 GCATGCTGCCTTGCTGATATTTTTAGAATTTATGCCCCAGAAGCTCCATATACATCGCCT 397

Qy      101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      398 GATAAACTGAAGGATATATTTATGTTCATAACAAGACAACTAAAGGGTCTTGAAGATACA 457

Qy      121 LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSer 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      458 AAGAGTCCACAGTTTAATAGATATTTCTATTTGCTAGAGAATATTGCATGGGTGAAATCC 517

Qy      141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
            |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db      518 TATAATATATGTTTTGAGCTGGAAGATTGCAATGAGATTTTTACGCAGCTATATAGAACA 577

Qy      161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      578 TTATTTTCGGTGATCAATAATGGACATAATCAGAAAGTTCATATGCATATGGTGGATTTA 637

Qy      181 MetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
            ||||||||||||:::|||||||||||||||||||||||||||||||||||:::||||||
Db      638 ATGAGCTCTATCGTTTGTGAAGGAGATACAGTTTCTCAGGAGCTGCTCGATAGTGTCCTA 697

Qy      201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
            |||||||||||||||||||||||||||      :::||||||||||||||||||||||||
Db      698 GTGAACTTGGTTCCTGCACACAAGAATTTTAAACAACAAGCTTATGACCTGGCAAAAGCT 757

Qy      221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
            ||||||||:::||||||||||||||||||||||||||||||||   ||||||:::||||||
Db      758 CTGCTTAAGAAAACTGCTCAAGCTATTGAACCATACATTACAAA-TTTTTCACCAAGTT 816

Qy      241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisVal-PheAspLeu-IleL 260
            |||:::||||||||||||||||||||||||||||||||||||    ::: |||||||| ||||
Db      817 CTCCTGCTTGGCAAAACCTCTATTAGTGATTTGTCAGAACCCCTGGTTCGATCTGGATTC 876

Qy      260 euGluLeuTyrAsnIleAspSerHis-LeuLeuLeuSerValLeuPro 275
            |||||||||||       ||||||||| ||||||       |||||||||
Db      877 TGGAGCTCTACACATTCGATAGCCATTTTGCTTCCTTTCGTTTTACCT 924

RESULT 15
BX314916
```

```
LOCUS       BX314916                 829 bp    mRNA    linear   EST 07-APR-2003
DEFINITION  BX314916 AGENAE Rainbow trout multi-tissues subtracted library
            (tcay) Oncorhynchus mykiss cDNA clone tcay0028b.k.14 5prim, mRNA
            sequence.
ACCESSION   BX314916
VERSION     BX314916.1  GI:29595561
KEYWORDS    EST.
SOURCE      Oncorhynchus mykiss (rainbow trout)
  ORGANISM  Oncorhynchus mykiss
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE   1  (bases 1 to 829)
  AUTHORS   Govoroun,M., Guiguen,Y. and Le Gac,F.
  TITLE     Construction and primary characterization of normalized cDNA
            libraries in rainbow trout, Oncorhynchus mykiss
  JOURNAL   Unpublished
COMMENT     Contact: Guiguen Y
            INRA - SCRIBE
            Campus de beaulieu, RENNES cedex, 35042, France
            Tel: 02.23.48.50.09
            Fax: 02.23.48.50.20
            Email: Yann.Guiguen@beaulieu.rennes.inra.fr
            Sequence cleaned of vector, adaptator and repetitions. Contact us
            at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
            sequence.
            Plate: 0028  row: k  column: 14
            Seq primer: M13F.
FEATURES             Location/Qualifiers
     source          1..829
                     /organism="Oncorhynchus mykiss"
                     /mol_type="mRNA"
                     /db_xref="taxon:8022"
                     /clone="tcay0028b.k.14"
                     /tissue_type="adipose tissue, blood, brain,
                     differentiating gonads, gills, interrenal, intestine,
                     kidney, liver, muscle, ovary, pituitary, testis"
                     /dev_stage="from embryos to adults"
                     /lab_host="DH10B"
                     /clone_lib="AGENAE Rainbow trout multi-tissues subtracted
                     library (tcay)"
                     /note="Vector: pT7T3D-pac; Clone distribution : AGENAE
                     Resource centre. Francois PIUMI,
                     Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
                     Etude du genome (LREG), Domaine de Vilvert, 78352,
                     Jouy-en-Josas cedex, FRANCE"
BASE COUNT      217 a    228 c    210 g    171 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:              8.11e-83      Length:        829
Score:                  1231.00       Matches:       234
Percent Similarity:     91.61%        Conservative:  17
Best Local Similarity:  85.40%        Mismatches:    22
Query Match:            17.11%        Indels:        1
DB:                     13            Gaps:          0

US-09-512-581B-2 (1-1391) x BX314916 (1-829)

Qy        854 LeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysPro 873
              |||         |||||||||||||||||||||||||||||||||||||:::   ||||||
Db          3 CTCGAGGCTATACTTCACAGTGATGGAGACCTGACAGAGCAGGGCAAGATGGGTAAACCA 62

Qy        874 AspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluPro 893
              ||||||||||||||||||||||||||          |||::::::||||||||||||||||
Db         63 GATATGTCCAGGCTGAGGTTGGCAGCAGCGTGTGCCCTACTGAAGCTGGCACAGGAGCCC 122

Qy        894 CysTyrHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAsp 913
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        123 TGCTACCATGAGATCATCACCCTGGAGCAGTACCAACTCTGTGCCCTGGTCATCAATGAT 182

Qy        914 GluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeu 933
              ||||||||||||||||||||   |||||||||||||||||||||||||||   ||||||
Db        183 GAGTGTTACCAGGTGCGCCAGGGCTTTGCCCAGAAGCTCCATAAGGGTCTGTGCCGTCTG 242

Qy        934 ArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGlu 953
              |||||||||||||||||:::|||:::   ||||||||||||||||||||||||||||||
Db        243 CGTCTGCCCCTGGAGTACCTGGCGGTGTTTGCCCTGTGTGCCAAGGACCCTGTGAAGGAG 302

Qy        954 ArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeu 973
              ||||||||||||||||||||||||||||||||||||||||      :::||||||||||||||
Db        303 AGGCGGGCCCACGCCAGGCAGTGTCTGGTGAAGAACATTACAATCAGGAGAGAGTACCTT 362

Qy        974 LysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValPro 993
              |||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db        363 AAACAACACGCTGCCGTCAGCGAGAAGTTGTTCTCTCTGCTCCCTGAGTATGTGGTCCCC 422

Qy        994 TyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGln 1013
              |||   ||||||||||||   ||||||||||||||:::||||||||||||||||||||||
Db        423 TATGCCATCCACCTGCTGGTCCATGACCCAGACTACATCAAGGTCCAGGACATTGAACAG 482

Qy       1014 LeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGlu 1033
              ||||||||:::||||||   |||||||||:::|||||||||:::|||:::||||||||
Db        483 CTCAAAGACATCAAGGAGTCTCTGTGGTTCATCCTGGAGATCATCATGTCTAAGAATGAG 542

Qy       1034 AsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAla 1053
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 AACAACAGCCATGCCTTCATCAGGAAGATGGTGGAGAACATCAAACAGACCAAGGATGCA 602

Qy       1054 GlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThr-ValCysAspValAlaMe 1073
              |||      |||   |||||||||:::||||||||||||||||||  ||||||||||||||||
Db        603 CAGGCTCCTACCGACGCCAAGATCAACGAGAAGTTGTACACGGGTGTGTGACGTGGCCAT 662

Qy       1073 tAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProValLe 1093
              ||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 GAACATCATCATCTCTAAGAGCACCACCTACAGTCTGGAGTCTCCCAAAGACCCTGTCCT 722

Qy       1093 uProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuPr 1113
              ||||   |||:::||||||:::|||   :::||||||   :::||||||||||||||||
Db        723 TCCCACACGCTACTTCACCAAGCCTGGACAGAATTTCAGNCACACAAAGAACTATCTCCC 782

Qy       1113 oProGluMetLysSerPhePheThrProGlyLysProLys 1126
              |   |||||||||||||||||||       |||||||||
Db        783 NTCAGAGATGAAGTCCTTCTTTACANCCTGGAAGCCTAAA 822

Search completed: September 25, 2003, 01:40:14
Job time : 6186 secs
```

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         September 25, 2003, 00:10:16 ; Search time 852 Seconds
                                           (without alignments)
                                           4064.415 Million cell updates/sec

Title:          US-09-512-581B-2
Perfect score:  7193
Sequence:       1 MAHSKTRTNDGKITYPPGVK..........QKGRGRPSKTPSPSQPKKNV 1391

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       1678620 seqs, 1244745471 residues

Total number of hits satisfying chosen parameters:      3357240

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09512581/runat_23092003_163549_5590/app_query.fasta_1.1543
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09512581_@CGN_1_1_474_@runat_23092003_163549_5590
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Published_Applications_NA:*
                1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
                2:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
                3:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
                4:  /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
                5:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
                6:  /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
                7:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                8:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
                9:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
                10:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
                11:  /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
                12:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                13:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
                14:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
                15:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
                16:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
                17:  /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1409.5	19.6	2496	9	US-09-822-849A-496	Sequence 496, App
	2	762.5	10.6	1217	9	US-09-925-302-357	Sequence 357, App
	3	698	9.7	439	13	US-10-040-739-107	Sequence 107, App
c	4	598	8.3	471	11	US-09-918-995-3514	Sequence 3514, Ap
c	5	349	4.9	530	10	US-09-879-536-682	Sequence 682, App
	6	276	3.8	413	11	US-09-918-995-20720	Sequence 20720, A
	7	276	3.8	433	11	US-09-918-995-33817	Sequence 33817, A
	8	232	3.2	8493	13	US-10-071-766-51	Sequence 51, Appl
	9	231.5	3.2	9274	10	US-09-885-535-3	Sequence 3, Appli
	10	231	3.2	8503	14	US-10-037-270-130	Sequence 130, App
	11	229	3.2	7992	11	US-09-893-519A-140	Sequence 140, App
	12	229	3.2	8491	12	US-10-133-013-260	Sequence 260, App
	13	228	3.2	7034	14	US-10-118-513A-11	Sequence 11, Appl
	14	225	3.1	5334	12	US-09-882-227-623	Sequence 623, App
	15	205.5	2.9	3727	14	US-10-171-581-49	Sequence 49, Appl
	16	205.5	2.9	7694	12	US-10-096-534-34	Sequence 34, Appl
	17	205.5	2.9	8063	12	US-09-814-353-21776	Sequence 21776, A
	18	205	2.8	7497	10	US-09-960-253-175	Sequence 175, App
	19	205	2.8	7792	13	US-10-044-090-359	Sequence 359, App
	20	200.5	2.8	4883	14	US-10-118-513A-5	Sequence 5, Appli
	21	200.5	2.8	7215	14	US-10-118-513A-13	Sequence 13, Appl
	22	200.5	2.8	7215	14	US-10-118-513A-15	Sequence 15, Appl
	23	200.5	2.8	10281	14	US-10-084-817-361	Sequence 361, App
	24	200	2.8	10281	12	US-10-240-965-201	Sequence 201, App
	25	197.5	2.7	10300	10	US-09-960-253-145	Sequence 145, App
	26	197	2.7	10190	10	US-09-864-864-292	Sequence 292, App
	27	197	2.7	10211	10	US-09-954-456-1153	Sequence 1153, Ap
	28	197	2.7	10211	10	US-09-967-768A-186	Sequence 186, App
	29	197	2.7	10211	11	US-09-918-624B-2	Sequence 2, Appli
	30	196	2.7	10096	10	US-09-960-253-163	Sequence 163, App
	31	194	2.7	5967	14	US-10-171-581-158	Sequence 158, App
	32	193	2.7	5612	14	US-10-128-714-1388	Sequence 1388, Ap
	33	193	2.7	5612	14	US-10-128-714-6388	Sequence 6388, Ap
	34	193	2.7	7611	14	US-10-128-714-388	Sequence 388, App
	35	193	2.7	7612	14	US-10-128-714-5388	Sequence 5388, Ap
	36	191	2.7	4455	11	US-09-998-027-7	Sequence 7, Appli
	37	191	2.7	10302	10	US-09-782-378A-23	Sequence 23, Appl
	38	191	2.7	16531	12	US-10-101-510-667	Sequence 667, App
	39	190.5	2.6	2212	9	US-09-925-297-142	Sequence 142, App
c	40	189.5	2.6	1639	12	US-10-017-161-2225	Sequence 2225, Ap
	41	189	2.6	5397	14	US-10-128-714-7388	Sequence 7388, Ap
	42	189	2.6	12337	14	US-10-171-311-5	Sequence 5, Appli
	43	189	2.6	12462	14	US-10-171-311-1	Sequence 1, Appli
	44	187.5	2.6	6452	14	US-10-205-823-418	Sequence 418, App
	45	186.5	2.6	5194	11	US-09-998-027-6	Sequence 6, Appli

ALIGNMENTS

```
RESULT 1
US-09-822-849A-496
; Sequence 496, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
;  APPLICANT: Wong, Gordon G.
;  APPLICANT:  Clark, Hilary
;  APPLICANT:  Fechtel, Kim
;  APPLICANT:  Agostino, Michael J.
;  APPLICANT:  Howes, Steven H.
;  APPLICANT:  Resnick, Richard J.
;  APPLICANT:  Gulukota, Kamalakar
;  APPLICANT:  Graham, James R.
;  APPLICANT:  Genetics Institute, Inc.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
;  FILE REFERENCE: GIN 6403
;  CURRENT APPLICATION NUMBER: US/09/822,849A
;  CURRENT FILING DATE:  2001-09-04
;  PRIOR APPLICATION NUMBER: 60/195,582
;  PRIOR FILING DATE: 2000-04-06
;  NUMBER OF SEQ ID NOS: 598
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 496
;   LENGTH: 2496
;   TYPE: DNA
```

```
;   ORGANISM: Homo sapiens
US-09-822-849A-496

Alignment Scores:
Pred. No.:                8.26e-118     Length:         2496
Score:                    1409.50       Matches:        300
Percent Similarity:       65.96%        Conservative:   76
Best Local Similarity:    52.63%        Mismatches:     123
Query Match:              19.60%        Indels:         72
DB:                       9             Gaps:           8

US-09-512-581B-2 (1-1391) x US-09-822-849A-496 (1-2496)

Qy     717 LysSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIle 736
           |||:::|||:::|||   |||   |||||||||   |||:::||||||||||||||||||
Db       4 AAAGCAAAGAGGGGTACTCCACACCAAGCAAAACAGGCTGTGCACTGTATACACGCCATA 63

Qy     737 PheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAsp 756
           |||::::::||||||   |||   |||||||||||||||
Db      64 TTCACAAATAAAGAAGTCCAGCTTGCACAGATTTTTGAGTCA------------------ 105

Qy     757 ProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeu 776

Db     105 ------------------------------------------------------------ 105

Qy     777 AlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAsp 796

Db     105 ------------------------------------------------------------ 105

Qy     797 LeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGlu 816
                                   |||:::|||   |||||||||   |||||||||
Db     106 ---------------------ACAGGTGAAAAGAATGGAAAACTGTGGTCTCCAGATGAA 144

Qy     817 GluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeu 836
           ||||||||||||||   :::   |||:::||||||||||||:::::::||||||||||||
Db     145 GAGGTTTCCCCTGAAGTACTAGCAAAGGTACAGGCAATTAAACTTCTGGTAAGGTGGCTG 204

Qy     837 LeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThr 856
           ||||||||||||||||||   |||||||||      ||||||||||||||||||:::
Db     205 TTGGGTATGAAAAACAACCAGTCTAAATCTGCCAATTCAACCCTTCGGTTATTATCAGCG 264

Qy     857 IleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSer 876
           :::|||   |||:::||||||||||||||||||   :::|||||||||   |||||||||
Db     265 ATGTTGGTTAGTGAGGGTGACCTGACAGAGCAAAAGAGGATCAGTAAATCTGATATGTCT 324

Qy     877 ArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHis 896
           ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db     325 CGCTTGCGATTAGCTGCTGGTAGTGCCATAATGAAGCTTGCTCAGGAACCTTGTTACCAT 384

Qy     897 GluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyr 916
           ||||||||||||   ||||||:::||||||||||||||||   ||||||||||||||||||
Db     385 GAAATTATTACCCCAGAACAGTTTCAGCTCTGTGCACTTGTTATTAATGATGAGTGTTAC 444

Qy     917 GlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuPro 936
           ||||||||||||:::|||||||||||||||||||||   |||   :::|||   ||||||
Db     445 CAAGTAAGGCAGATATTTGCTCAGAAGCTGCATAAGGCACTTGTGAAGTTACTGCTCCCA 504

Qy     937 LeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAla 956
           ||||||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db     505 TTGGAGTATATGGCGATCTTTGCCTTGTGTGCCAAAGATCCTGTGAAGGAGAGAAGAGCA 564

Qy     957 HisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHis 976
           |||||||||||||||||||||:::|||||||||:::::||||||||||||:::||||||:::
Db     565 CACGCACGACAATGTTTACTGAAAAATATCAGTATACGCAGGGAATACATTAAGCAGAAT 624

Qy     977 AlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIle 996
           :::||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db     625 CCTATGGCTACTGAGAAATTATTATCACTGTTGCCTGAATATGTAGTTCCATACATGATT 684

Qy     997 HisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAsp 1016
           |||||||||||||||||||||||||:::   :::   ||||||:::::||||||:::|||
Db     685 CACCTGCTAGCCCATGATCCAGATTTTACAAGATCACAAGATGTTGATCAGCTTCGTGAT 744

Qy    1017 ValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSer 1036
           :::||||||||||||||||||:::||||||:::||||||   ||||||||||||||||||
Db     745 ATCAAAGAGTGCCTATGGTTCATGCTTGAAGTTTTAATGACAAAGAATGAAAACAATAGC 804

Qy    1037 HisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPro 1056
           |||||||||::::::|||||||   ||||||||||||   |||:::|||||||||   |||
Db     805 CATGCCTTTATGAAGAAGATGGCAGAGAACATCAAGTTAACCAGAGATGCCCAGTCTCCA 864

Qy    1057 AspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIle 1076
           |||::::::|||   |||||||||||||||||||||||||||||||||:::   :::|||
Db     865 GATGAATCCAAGACAATGAAAAACTGTATACAGTATGTGATGTGGCTCTCTGTGTTATA 924

Qy    1077 MetSerLysSerThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArg 1096
              |||||||||         :::::::::   ||||||   |||   |||
Db     925 AATAGTAAAAGTGCTTTGT--GCAATGCAGATTCCCAAAGGACCCAGCCTCCAATGAAAT 982

Qy    1097 PhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMet 1116
           |||:::|||   |||:::|||:::|||   |||   |||:::|||:::      |||
Db     983 TTTTACACA---CCTGAAAAGGACTTCTGTAACGATAAGAGTTATATTTCAGAAGAGACA 1039

Qy    1117 LysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLys 1136
           :::               ||||||||||||         ||||||||||||||||||||||
Db    1040 AGAGTACTTCTGTTAACAGGAAAGCCAAAGCCTGCTGGAGTACTAGGTGCAGTAAATAAG 1099

Qy    1137 ProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSer 1156
           |||||||||:::   |||::::::         :::|||:::      ||||||   |||
Db    1100 CCTTTATCAGCAACGGGAAGGAAACCCTATGTTAGAAGCACTGGCACTGAGACTGGAAGC 1159

Qy    1157 ------AsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArg--- 1173
                 |||   :::|||   |||||||||:::   |||   :::   :::
Db    1160 AATATTAATGTAAATTCAGAGCTGAACCCTTCAACCGGAAATCGATCAAGGGAACAGAGT 1219

Qy    1174 LeuAspSerSerGluMetAspHisSerGluAsnGluAsp------TyrThrMetSerSer 1191
           :::::::::|||         ||||||||||||:::            :::   |||
Db    1220 TCAGAGGCAGCAGAAACTGGAGTTAGTGAAAATGAAGAGAACCCTGTGAGGATTATTTCA 1279

Qy    1192 ProLeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeu 1211
              |||   |||   |||               :::      |||:::      |||
Db    1280 GTCACACCTGTAAAGAATATTGACCCAGTAAAGAATAAG------GTAAAAATGCATTTG 1333

Qy    1212 GluLysProArgGlyArgLysLysThrProValThrGluGlnGlu-GluLysLeuGlyMe 1231
           ::::::   :::   :::   |||                      :::|||||||||
Db    1334 CAAAGGGAGAAAATGAAGGCCAAACAGAAGCAGGCTCCAGCTTCTGCAAAAACTTGGATT 1393

Qy    1231 tAspAspLeu------------ThrLysLeuValGlnGluGln---------------Ly 1242
               ::::::              ||||||   |||                      ||
Db    1394 CACAAATGTCCCTGAACAGAAAATGAAGCTCACTTCAGAACACACACTCTCTGCCTTGAA 1453

Qy    1242 sProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluGl 1262
            |   |||   :::      |||            |||   :::|||:::   :::   ||
Db    1454 AACTAAAGAGACTATTACTTCCTTTTCACATGACCACAAGTCCTCTGATGGAAATGTACA 1513

Qy    1262 nGlnTrpProGluGluLysArgLeuLys 1271
            ||||         :::|||||||||
Db    1514 GCAGAAACTCTTGAGAGAGAGGCTAAAA 1541

RESULT 2
US-09-925-302-357
; Sequence 357, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA104
;  CURRENT APPLICATION NUMBER: US/09/925,302
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05918
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 896
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 357
```

```
;    LENGTH: 1217
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc feature
;    LOCATION: (1141)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: misc feature
;    LOCATION: (1149)
;    OTHER INFORMATION: n equals a,t,g, or c
;    NAME/KEY: misc feature
;    LOCATION: (1157)
;    OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-357

Alignment Scores:
Pred. No.:              1.76e-59      Length:         1217
Score:                  762.50        Matches:        185
Percent Similarity:     56.88%        Conservative:   63
Best Local Similarity:  42.43%        Mismatches:     121
Query Match:            10.60%        Indels:         68
DB:                     9             Gaps:           15

US-09-512-581B-2 (1-1391) x US-09-925-302-357 (1-1217)

Qy        965 AsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeu 984
              ||||||:::::|||||||||||:::|||||:::           :::|||||||||||
Db          2 AATATCAGTATACGCAGGGAATACATTAAGCAGAATCCTATGGCTACTGAGAAATTATTA 61

Qy        985 SerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAsp 1004
              |||||||||||||||||||||||||||||     |||||||||||||||||||||||||
Db         62 TCACTGTTGCCTGAATATGTAGTTCCATACATGATTCACCTGCTAGCCCATGATCCAGAT 121

Qy       1005 TyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheVal 1024
              :::   :::   |||||::::::||||||:::|||:::|||||||||||||||||:::
Db        122 TTTACAAGATCACAAGATGTTGATCAGCTTCGTGATATCAAAGAGTGCCTATGGTTCATG 181

Qy       1025 LeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetVal 1044
              ||||||:::||||||   |||||||||||||||||||||||||||::::::||||||
Db        182 CTTGAAGTTTTAATGACAAAGAATGAAAACAATAGCCATGCCTTTATGAAGAAGATGGCA 241

Qy       1045 GluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLys 1064
              ||||||||||||   |||:::|||||||||   ||||||::::::|||   |||||||||
Db        242 GAGAACATCAAGTTAACCAGAGATGCCCAGTCTCCAGATGAATCCAAGACAAATGAAAAA 301

Qy       1065 LeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSer 1084
              ||||||||||||||||||||||||:::    :::|||   |||||||||         :::
Db        302 CTGTATACAGTATGTGATGTGGCTCTCTGTGTTATAAATAGTAAAAGTGCTTTGTGCAAT 361

Qy       1085 LeuGluSerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsn 1104
                 :::||||||||||||||||||||||||   :::||||||||||||||||:::|||:::
Db        362 GCAGATTCACCAAAGGACCCAGTCCTCCCAATGAAATTTTTTACACAACCTGAAAAGGAC 421

Qy       1105 PheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePheThrProGlyLys 1124
              |||   |||   |||:::|||:::         |||   :::               ||||||
Db        422 TTCTGTAACGATAAGAGTTATATTTCAGAAGAGACAAGAGTACTTCTGTTAACAGGAAAG 481

Qy       1125 ProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGln 1144
              ||||||         ||||||||||||||||||||||||||||||:::   |||::::::
Db        482 CCAAAGCCTGCTGGAGTACTAGGTGCAGTAAATAAGCCTTTATCAGCAACGGGAAGGAAA 541

Qy       1145 SerGlnThrLysSerSerArgMetGluThrValSer------AsnAlaSerSerSerSer 1162
                    :::|||:::         ||||||   |||         |||   :::|||
Db        542 CCCTATGTTAGAAGCACTGGCACTGAGACTGGAAGCAATATTAATGTAAATTCAGAGCTG 601

Qy       1163 AsnProSerSerProGlyArgIleLysGlyArg---LeuAspSerSerGluMetAspHis 1181
              ||||||||||:::         |||   :::   :::         :::::::::|||
Db        602 AACCCTTCAACCGGAAATCGATCAAGGGAACAGAGTTCAGAGGCAGCAGAAACTGGAGTT 661

Qy       1182 SerGluAsnGluAsp------TyrThrMetSerSerProLeuProGlyLysLysSerAsp 1199
              ||||||||||||:::            :::   |||      |||   |||         |||
Db        662 AGTGAAAATGAAGAGAACCCTGTGAGGATTATTTCAGTCACACCTGTAAAGAATATTGAC 721

Qy       1200 LysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLys 1219

Db        721 ------------------------------------------------------------ 721

Qy       1220 ThrProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuValGln 1239
                 ||||||         :::|||      :::      |||      |||
Db        722 ---CCAGTAAAGAATAAGGAA------ATTAATTCTGATCAGGCTACC------------ 760

Qy       1240 GluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSer 1259
                 |||-            |||   |||   :::||||||         ||||||:::   :::
Db        761 ---CAGGGCAACATCAGCAGTGACCGAGGAAAGAAAAGAACAGTAACAGCAGCTGGTGCA 817

Qy       1260 AspGlu---GlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGlu 1278
              :::         ||||||      :::||||||
Db        818 GAGAATATCCAACAAAAAACAGATGAGAAA---------------------------GTA 850

Qy       1279 AspGluGlnAsnSerPro---ProLysLysGlyLysArgGlyArgProProLysProLeu 1297
              ||||||            |||   |||   |||   :::|||||||||   ||||||
Db        851 GATGAATCGGGACCTCCCGCCCCTTCCAAACCCAGGAGAGGACGTCGACCCAAGTCTGAA 910

Qy       1298 Gly---GlyGlyThrProLysGluGluProThrMetLysThrSerLysLysGlySerLys 1316
                    |||         |||   :::               |||         ||||||   |||
Db        911 TCTCAGGGCAATGCTACCAAAAATGATGATCTAAATAAACCTATTAACAAGGGAAGGAAG 970

Qy       1317 LysLysSer---GlyProProAlaProGluGluGluGlu--------------------- 1328
              :::   :::      |||      :::|||         |||
Db        971 AGAGCTGCAGTGGGTCAGGAGAGCCCTGGGGGTTTGGAAGCAGGTAATGCCAAAGCACCC 1030

Qy       1329 ------GluGluGluArgGlnSerGlyAsnThrGluGlnLys------------SerLys 1342
                    ::::::      |||   |||:::|||      :::|||               :::|||
Db       1031 AAACTGCAAGATTT-AGCCAAAAAGGCAGCACCAGCAGAAAGACMAATTGACTTACAAAG 1089

Qy       1343 SerLysGlnHis------ArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
                 |||   |||         |||         :::|||:::|||:::
Db       1090 GTRAAAATGCATTTGCAAAGGGAGAAAATGAAGGCCAAACAGAAGCAGGCTNCCAGYTTN 1149

Qy       1361 SerSerAlaIle------------------GluSerThrGlnSerThr 1370
                    :::                     |||   |||   ||||||
Db       1150 TGCAAAANCTTGGATTACAATGKCCTGACAGAAATGACTTATTCAACC 1197

RESULT 3
US-10-040-739-107
; Sequence 107, Application US/10040739
; Publication No. US20020173635A1
;  GENERAL INFORMATION:
;       APPLICANT: Jacobs, Kenneth
;                  McCoy, John
;                  LaVallie, Edward
;                  Racie, Lisa
;                  Merberg, David
;                  Treacy, Maurice
;                  Spaulding, Vikki
;       TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
;       NUMBER OF SEQUENCES: 1519
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Genetics Institute, Inc.
;            STREET: 87 CambridgePark Drive
;            CITY: Cambridge
;            STATE: Massachusetts
;            COUNTRY: U.S.A
;            ZIP: 02140
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy Disk
;            COMPUTER: IBM PC Compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/040,739
;            FILING DATE: 07-Jan-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/036,520
;            FILING DATE: 03-JUN-1998
```

```
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Brown, Scott A.
;            REGISTRATION NUMBER: 32,724
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (617) 498-8224
;            TELEFAX: (617) 876-5851
;   INFORMATION FOR SEQ ID NO: 107:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 439 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 107:
US-10-040-739-107

Alignment Scores:
Pred. No.:               2.64e-54        Length:        439
Score:                   698.00          Matches:       137
Percent Similarity:      97.86%          Conservative:  0
Best Local Similarity:   97.86%          Mismatches:    3
Query Match:             9.70%           Indels:        0
DB:                      13              Gaps:          0

US-09-512-581B-2 (1-1391) x US-10-040-739-107 (1-439)

Qy       1073 MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProVal 1092
              |||      |||||||||||||||||||||||||||||||||||||||||||||||||||
Db         14 ATGGCCTACATCATGTCAAAGAGTACTACATACAGTTTGGAATCTCCTAAAGACCCGGTA 73

Qy       1093 LeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeu 1112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         74 CTACCAGCTCGTTTCTTCACTCAACCTGACAAGAATTTCAGTAACACCAAAAATTATCTG 133

Qy       1113 ProProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGly 1132
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        134 CCTCCTGAAATGAAATCATTTTTCACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGA 193

Qy       1133 AlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMet 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 GCTGTTAACAAGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAAATCATCACGAATG 253

Qy       1153 GluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGly 1172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        254 GAAACTGTAAGCAATGCAAGCAGCAGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGG 313

Qy       1173 ArgLeuAspSerSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSerPro 1192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        314 AGGCTTGATAGTTCTGAAATGGATCACAGTGAAAATGAAGATTACACAATGTCTTCACCT 373

Qy       1193 LeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGlu 1212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        374 TTGCCGGGGAAAAAAAGTGACAAGAGAGACGACTCTGATCTTGTAAGGTCTGAATCGGAG 433

RESULT 4
US-09-918-995-3514/c
; Sequence 3514, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3514
;   LENGTH: 471
;   TYPE: DNA
;   ORGANISM: Homo sapiens
```

```
US-09-918-995-3514

Alignment Scores:
Pred. No.:               3.62e-45        Length:        471
Score:                   598.00          Matches:       110
Percent Similarity:      88.89%          Conservative:  18
Best Local Similarity:   76.39%          Mismatches:    16
Query Match:             8.31%           Indels:        0
DB:                      11              Gaps:          0

US-09-512-581B-2 (1-1391) x US-09-918-995-3514 (1-471)

Qy        437 LysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGlu 456
              |||   |||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db        459 AAAGGCAAACTTCTGCATATTTATTATCAGAACAGCATTGACGACAAACTGTTGGTAGAG 400

Qy        457 ArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCys 476
              :::||||||||||||:::||||||||||||||||||||||||   |||||||||||||||
Db        399 AAAATCTTTGCTCAGTATCTTGTCCCCCACAACCTGGAAACAGAAGAGAGAATGAAATGC 340

Qy        477 LeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrp 496
              ||||||||||||||||||:::||||||   ||||||||||||||||||||||||||||||
Db        339 TTATATTACTTATATGCTAGTTTGGATCCAAATGCTGTAAAAGCTCTCAACGAAATGTGG 280

Qy        497 LysCysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnPro 516
              ||||||||||||:::||||||      |||:::::::||||||||||||   |||||||||
Db        279 AAGTGTCAGAACATGCTTCGGAGCCATGTACGCGAACTATTGGATTTGCACAAGCAGCCT 220

Qy        517 LysThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeu 536
                 ::::::|||:::      |||:::|||   |||:::|||   |||   :::||||||
Db        219 ACATCAGAGGCTAACTGTTCTGCCATGTTTGGAAAACTGATGACCATAGCAAAGAATTTG 160

Qy        537 ProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAsp 556
              |||||||||||||||||||||||||||:::|||||||||   |||||||||   ||||||
Db        159 CCTGACCCCGGGAAAGCACAAGATTTTGTGAAGAAATTTAACCAGGTTCTCGGCGATGAT 100

Qy        557 GluLysIleArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAla 576
              ||||||:::|||   ||||||||:::|||:::|||||||||||||||||||||||||||||
Db         99 GAGAAACTTCGGTCTCAGTTGGAGTTATTAATTAGCCCAACCTGTTCTTGCAAACAAGCA 40

Qy        577 GluGlyCysVal 580
              :::   ||||||
Db         39 GATATTTGTGTG 28

RESULT 5
US-09-879-536-682/c
; Sequence 682, Application US/09879536
; Patent No. US20020144298A1
; GENERAL INFORMATION:
;  APPLICANT: Endege, Wilson O.
;  APPLICANT:  Steinmann, Kathleen E.
;  APPLICANT:  Astle, Jon H.
;  APPLICANT:  Burgess, Christopher C.
;  APPLICANT:  Bushnell, Steven E.
;  APPLICANT:  Carroll III, Eddie
;  APPLICANT:  Catino, Theodore J.
;  APPLICANT:  Derti, Adnan
;  APPLICANT:  Ford, Donna M.
;  APPLICANT:  Lewis, Marcia E.
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Schlegel, Robert
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION:  PRODUCTS
;  FILE REFERENCE: CCD-257 (US)
;  CURRENT APPLICATION NUMBER: US/09/879,536
;  CURRENT FILING DATE:  2001-09-21
;  PRIOR APPLICATION NUMBER: US 60/088,801
;  PRIOR FILING DATE: 1998-06-10
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 682
;   LENGTH: 530
;   TYPE: DNA
```

```
;   ORGANISM: Homo sapiens
US-09-879-536-682

Alignment Scores:
Pred. No.:              1.87e-22        Length:        530
Score:                  349.00          Matches:       67
Percent Similarity:     98.53%          Conservative:  0
Best Local Similarity:  98.53%          Mismatches:    1
Query Match:            4.85%           Indels:        0
DB:                     10              Gaps:          0

US-09-512-581B-2 (1-1391) x US-09-879-536-682 (1-530)

Qy         37 MetValValLysThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyr 56
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 ATGGTTGTGAAAACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTAT 154

Qy         57 LeuAsnLeuAlaLeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspVal 76
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        153 TTAAACCTAGCTTTACATCTTGCTTCAGATTTTTTCCTCAAGCATCCTGATAAGGATGTT 94

Qy         77 ArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaPro 96
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         93 CGCTTACTGGTAGCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCT 34

Qy         97 TyrThrSerProAspLysLeuLys 104
              ||||||||||||||||||||||||
Db         33 TACACATCCCCTGATAAACTAAAG 10

RESULT 6
US-09-918-995-20720
; Sequence 20720, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 20720
;   LENGTH: 413
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-918-995-20720

Alignment Scores:
Pred. No.:              5.49e-16        Length:        413
Score:                  276.00          Matches:       62
Percent Similarity:     52.94%          Conservative:  10
Best Local Similarity:  45.59%          Mismatches:    7
Query Match:            3.84%           Indels:        57
DB:                     11              Gaps:          1

US-09-512-581B-2 (1-1391) x US-09-918-995-20720 (1-413)

Qy         11 GlyLysIleThrTyrProProGlyValLysGluIleSerAspLysIleSerLysGluGlu 30
              |||         ||||||||||||||||||||||||:::|||||||||:::   :::|||
Db          7 GGCACGAGCGCTTACCCTCCGGGGGTAAAAGAGATCACCGACAAGATCACCACGGACGAG 66

Qy         31 MetValArgArgLeuLys------------------------------------------ 36
              |||::::::|||||||||
Db         67 ATGATCAAACGCCTGAAGGGGAAAACCAACTGCCTTTTCACTTGAGCTGTCTAGAAAGGC 126

Qy         36 ------------------------------------------------------------ 36

Db        127 CCATGTAGTGTGGAACTGAAGCCTTTCAACAACAGCAAGTGTCAACTTATAACCGTAAGA 186

Qy         36 ------------------------------------------------------------ 36

Db        187 ATGAGTTACCTTCAAAACACACCCTCCACGCCCAGTCAAGCCTTCAGATAACTGCAGCCT 246

Qy         37 -------MetValValLysThrPheMetAspMetAspGlnAspSerGluGluGluLysGl 54
                     ||||||||||||||||||||||||||||||||||||||||||:::||||||::
Db        247 TGACCAGATGGTAGTGAAAACCTTTATGGATATGGATCAGGACTCAGAAGATGAAAAACA 306

Qy         54 uLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLy 74
               :   ||||||   ||||||||||||||||||||:::||||||||||::::::|||   ||
Db        307 GCAGTATCTCCCACTAGCCTTGCATCTTGCATCTGAATTCTTCCTCAGGAACCCCAATAA 366

Qy         74 sAspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArg 89
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        367 AGATGTGCGTCTCCTTGTAGCATGTTGTTTGGCTGATATCTTTCGT 412

RESULT 7
US-09-918-995-33817
; Sequence 33817, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33817
;   LENGTH: 433
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-918-995-33817

Alignment Scores:
Pred. No.:              5.91e-16        Length:        433
Score:                  276.00          Matches:       55
Percent Similarity:     100.00%         Conservative:  0
Best Local Similarity:  100.00%         Mismatches:    0
Query Match:            3.84%           Indels:        0
DB:                     11              Gaps:          0

US-09-512-581B-2 (1-1391) x US-09-918-995-33817 (1-433)

Qy       1337 ThrGluGlnLysSerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAla 1356
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 ACGGAACAGAAGTCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCA 62

Qy       1357 GluSerProGluSerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGly 1376
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 GAATCTCCTGAATCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGA 122

Qy       1377 ArgProSerLysThrProSerProSerGlnProLysLysAsnVal 1391
              |||||||||||||||||||||||||||||||||||||||||||||
Db        123 AGACCATCAAAAACGCCATCACCATCACAACCAAAAAAAAATGTC 167

RESULT 8
US-10-071-766-51
; Sequence 51, Application US/10071766
; Publication No. US20020192678A1
; GENERAL INFORMATION:
;  APPLICANT: Huei-Mei Chen
;  TITLE OF INVENTION: GENES EXPRESSED IN SENESCENCE
;  FILE REFERENCE: PA-0043 US
;  CURRENT APPLICATION NUMBER: US/10/071,766
;  CURRENT FILING DATE:  2002-02-07
;  NUMBER OF SEQ ID NOS: 144
;  SOFTWARE: PERL Program
; SEQ ID NO 51
;   LENGTH: 8493
;   TYPE: DNA
```

```
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    OTHER INFORMATION: Incyte ID No. US20020192678A1 441283.3
US-10-071-766-51

Alignment Scores:
Pred. No.:              5.96e-10       Length:         8493
Score:                  232.00         Matches:        285
Percent Similarity:     35.41%         Conservative:   267
Best Local Similarity:  18.28%         Mismatches:     541
Query Match:            3.23%          Indels:         467
DB:                     13             Gaps:           67

US-09-512-581B-2 (1-1391) x US-10-071-766-51 (1-8493)

Qy        17 ProGlyValLysGluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLys 36
             |||   |||   |||:::|||            :::   |||   ::::::::|||   :::
Db      1087 CCTTATGTTAATGAGGTATCA---------ACTGATGAAGCTCTCCTGAAAAGGTATAGA 1137

Qy        37 MetValValLysThrPheMetAspMetAspGlnAspSerGluGlu--------------- 51
                      |||   ||||||:::   :::   ||||||
Db      1138 ---------AAAGAAATAATGGATCTTAAAAAACAATTAGAGGAGGTTTCTTTAGAGACG 1188

Qy        52 ---------------GluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
                            ||||||:::            |||                  |||
Db      1189 CGGGCTCAGGCAATGGAAAAAGACCAATTGGCCCAACTTTTGGAAGAA------AAAGAT 1242

Qy        67 PhePheLeuLysHisProGlyLysAspVal------------ArgLeuLeuValAlaCys 82
                      |||         :::   :::            |||:::||||||
Db      1243 TTGCTTCAGAAAGTACAGAATGAGAAAATTGAAAACTTAACACGGATGCTGGTG------ 1296

Qy        83 CysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLys 102
                                                         ||||||
Db      1297 ---------------------------------------------ACCTCTTCTTCC--- 1308

Qy       103 LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLysSer 122
                                  :::|||   |||            ::::::   |||:::
Db      1309 --------------------CTCACGTTGCAA------------CAGGAATTAAAGGCT 1335

Qy       123 ProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsn 142
                :::      |||      :::   |||      |||      :::|||:::   |||
Db      1336 AAAAGAAAACGAAGAGTTACTTGGTGCCTTGGCAAAATTAACAAAATGAAGAACTCAAAC 1395

Qy       143 IleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPhe 162
                                                         |||         |||
Db      1396 ---------------------------------------------TATGCAGATCAATTT 1410

Qy       163 SerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSer 182
             ::::::         |||   |||      |||   |||         :::::|||:::
Db      1411 AATATACCAACAAAT---ATAACAACAAAAACACATAAGCTTTCTATAAATTTATTACGA 1467

Qy       183 SerIle---------IleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrVal 199
                |||         :::|||            |||         |||||||||:::
Db      1468 GAAATTGATGAATCTGTCTGTTCAGAGTCTGATGTTTTCAGTAACACTCTTGATACATTA 1527

Qy       200 Leu---ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAla 218
                   :::         ||||||   |||   ||||||::::::      :::
Db      1528 AGTGAGATAGAATGGAATCCAGCAACAAAGCTACTAAATCAGGAGAATATAGAA------ 1581

Qy       219 LysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsn 238

Db      1581 ------------------------------------------------------------ 1581

Qy       239 GlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeu 258
                                  :::   ::::::   |||               :::|||
Db      1582 ---------------------AGTGAGTTGAACTCACTTCGTGCTGACTATGATAATCTG 1620

Qy       259 IleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGlu 278
             :::|||               |||      |||   :::         ::::::|||
Db      1621 GTATTA---------------GACTATGAACAACTACGAACAGAAAAAGAAGAAATGGAA 1665

Qy       279 PheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLys 298
                |||||||||         ::::::   :::      :::   ::::::         |||
Db      1666 TTGAAATTAAAAGAAAAGAATGATTTGGATGAATTTGAGGCTCTAGAAAGAAAAACTAAA 1725

Qy       299 MetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyr 318
                         ||||||   |||:::
Db      1726 ------------AAAGATCAAGAGATGCAA------------------------------ 1743

Qy       319 LeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSer 338

Db      1743 ------------------------------------------------------------ 1743

Qy       339 HisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArg 358
                   |||::::::                  |||:::|||   ::::::   :::|||
Db      1744 ------CTAATTCATGAAATTTCGAACTTAAAGAATTTAGTTAAGCATCGAGAAGTATAT 1797

Qy       359 SerHisAspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAla 378
             :::   |||   |||      :::                              ::::::
Db      1798 AATCAAGATCTTGAGAATGAACTC------------------------------AGTTCA 1827

Qy       379 LysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeu 398
             |||   ::::::|||         :::      :::      ::::::|||
Db      1828 AAAGTAGAGCTGCTTAGAGAAAAGGAAGACCAGATTAAGAAGCTACAGGAATACATAGAC 1887

Qy       399 AspLysArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLys 418
             ::::::            :::      |||   |||:::
Db      1888 TCTCAAAAG---------CTAGAAAATATAAAAATGGACTTGTCA--------------- 1923

Qy       419 TyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAsp 438
             |||:::|||:::|||                  ||||||      :::         :::
Db      1924 TACTCATTGGAAAGCATTGAAGACCCAAAACAAATGAAGCAGACTCTGTTTGATGCTGAA 1983

Qy       439 LysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIle 458
                :::                                       |||      :::
Db      1984 ACTGTA------------------------------------GCCCTTGATGCCAAGAGA 2007

Qy       459 PheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyr 478
                :::   ::::::            |||||||||      |||:::||||||   |||
Db      2008 GAATCAGCCTTTCTTAGAAGTGAAAATCTGGAGTTGAAGGAGAAAATGAAAGAACTTGCA 2067

Qy       479 TyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCys 498
                |||      ::::::                              |||::::::
Db      2068 ACTACATACAAGCAAATGGAA---------------------AATGATATTCAGTTATAT 2106

Qy       499 GlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThr 518
             |||:::      |||      :::   |||      :::||||||   |||:::
Db      2107 CAAAGCCAATTGGAGGCAAAAAAGAAAATGCAAGTTGATCTGGAGAAAGAA--------- 2157

Qy       519 AspAlaSerValLysAlaIlePheSerLysValMetValIleThrArg------------ 534
                      :::::::::   |||:::::::::      :::|||
Db      2158 ---------TTACAATCTGCTTTTAATGAGATAACAAAACTCACCTCCCTTATAGATGGC 2208

Qy       535 ------------------------AsnLeuProAspProGlyLysAlaGlnAspPheMet 546
                                     ||||||         ||||||      |||
Db      2209 AAAGTTCCAAAAGATTTGCTCTGTAATTTGGAATTGGAAGGAAAGATTACTGATCTTCAG 2268

Qy       547 LysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeu 566
             |||:::      :::   :::|||::::::|||      :::|||:::::::::   :::|||
Db      2269 AAAGAACTAAATAAAGAAGTTGAAGAAAATGAAGCTTTGCGGGAAGAAGTC---ATTTTG 2325

Qy       567 ValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLys 586
             :::|||         |||               |||      :::   |||:::
Db      2326 CTTTCA---------GAATTGAAATCTTTACCTTCTGAAGTAGAAAGGCTGAGGAAAGAG 2376

Qy       587 LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGlu 606
             :::         |||   :::      |||   :::|||                  :::
Db      2377 ATA-----------CAAGACAAATCTGAAGAGCTCCATATAATAACATCAGAAAAAGAT 2424

Qy       607 ArgIleAlaProValHisIleAspThrGluSer---IleSerAlaLeuIleLysGlnVal 625
             ::::::            :::      ||||||   :::      |||::::::::::::
Db      2425 AAATTGTTTTCTGAAGTAGTTCATAAGGAGAGTAGAGTTCAAGGTTTACTTGAAGAAATT 2484

Qy       626 AsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAla 645
                |||:::   |||      |||         :::                  |||||||||
```

```
Db   2485 GGGAAAACAAAAGATGACCTAGCAACTACACAGTCGAATTATAAAAGCACTGATCAAGAA 2544

Qy    646 Ile--------------------------------------------------------- 646

Db   2545 TTCCAAAATTTCAAAACCCTTCATATGGACTTTGAGCAAAAGTATAAGATGGTCCTTGAG 2604

Qy    647 ---------ArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSer 663
                   |||         |||:::::::   :::
Db   2605 GAGAATGAGAGAATGAATCAGGAAATAGTTAATCTCTCTAAA------------------ 2646

Qy    664 PheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp--------- 680
                   |||:::   |||:::|||   |||     ||||||    :::
Db   2647 ------GAAGCCCAAAAATTTGATTCGAGTTTGGGTGCTTTGAAGACCGAGCTTTCTTAC 2700

Qy    681 --------------AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThr 695
                        ||||||    |||         ::::::     |||
Db   2701 AAGACCCAAGAACTTCAGGAGAAAACACGTGAGGTTCAAGAAAGACTA------AATGAG 2754

Qy    696 GlySerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHis 715
                :::::::::|||         :::      |||   |||     |||
Db   2755 ATGGAACAGCTGAAGGAACAATTAGAAAATAGAGATTCTCCGCTGCAAACTGTAGAAAGG 2814

Qy    716 HisLysSer--------------------------------------------------- 718
            |||:::
Db   2815 GAGAAAACACTGATTACTGAGAAACTGCAGCAAACTTTAGAAGAAGTAAAAACTTTAACT 2874

Qy    719 ---LysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePhe 737
             :::|||        :::|||   :::   ::::::
Db   2875 CAAGAAAAAGATGATCTAAAACAACTCCAAGAAAGCTTGCAAATT--------------- 2919

Qy    738 SerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro 757
                ::::::    :::   |||   :::   ::::::   ::::::|||
Db   2920 ---GAGAGGGACCAACTCAAAAGTGATATTCACGATACTGTTAACATGAATATAGATACT 2976

Qy    758 SerAsn---------------LeuGluHisLeuIleThrProLeuValThrIleGlyHis 772
                               ||||||   |||             ||||||
Db   2977 CAAGAACAATTACGAAATGCTCTTGAGTCTCTGAAACAACATCAAGAAACAATTAATACA 3036

Qy    773 IleAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPhe 792
          :::                                   ||||||    ::::::
Db   3037 CTA-----------------------------------AAATCGAAAATTTCTGAGGAA 3060

Qy    793 IleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrp 812
          :::   ::::::|||   |||
Db   3061 GTTTCCAGGAATTTGCATATG--------------------------------------- 3081

Qy    813 ValProAspGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMet 832
                   ||||||   :::   ||||||      :::   |||
Db   3082 ---------GAGGAAAATACAGGAGAAACTAAAGATGAATTTCAG--------------- 3117

Qy    833 ValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg 852
             :::   ::::::|||:::                     :::|||      :::
Db   3118 ---CAAAAGATGGTTGGCATAGATAAAAAACAGGATTTGGAAGCTAAAAATACC---CAA 3171

Qy    853 LeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLys 872
             ||||||       :::      :::      :::   :::|||   ||||||
Db   3172 ACACTAACTGCAGATGTTAAGGATAATGAGATAATTGAGCAACAAAGGAAGATATTTTCT 3231

Qy    873 ------ProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAla 890
                :::   :::   |||:::               |||   |||   :::
Db   3232 TTAATACAGGAGAAAAATGAACTCCAACAAATGTTAGAGAGTGTTATAGCAGAAAAGGAA 3291

Qy    891 Gln------------------GluProCysTyrHisGluIleIleThrLeuGluGlnTyr 904
          |||                  |||                  |||      |||:::
Db   3292 CAATTGAAGACTGACCTAAAGGAAAATATTGAAATGACCATTGAAAACCAGGAAGAATTA 3351

Qy    905 GlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGln 924
          :::|||               :::   ||||||      :::   :::::::::   ||||||
Db   3352 AGACTT------------CTTGGGGATGAACTTAAAAGCAACAAGAGATAGTTGCACAA 3399

Qy    925 LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAla 944
          :::   :::         :::   :::                  |||         |||
Db   3400 GAAAAGAACCATGCCATAAAGAAA---------GAAGGAGAGCTTTCTAGGACCTGTGAC 3450

Qy    945 LeuCysAla------LysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeu 962
             |||          ::::::   ::::||||||:::              ::::::
Db   3451 AGACTGGCAGAAGTTGAAGAAAAACTAAAGGAAAAGAGCCAGCAACTCCAAGAAAAACAG 3510

Qy    963 ValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGlu--- 981
          :::   :::|||||||:::   |||         :::                ::::::|||
Db   3511 CAACAACTTCTTAATGTACAAGAAGAGATGAGTGAGATGCAGAAAAAGATTAATGAAATA 3570

Qy    982 --------------------LysLeuLeuSerLeuLeuProGluTyrValValProTyr 994
                              |||         |||      |||::::::
Db   3571 GAGAATTTAAAGAATGAATTAAAGAACAAAGAATTGACATTGGAACATATGGAAACAGAG 3630

Qy    995 ThrIleHisLeuLeuAlaHis---AspProAspTyrValLysValGlnAspIle---Glu 1012
             :::   |||            :::   :::|||   :::|||:::   |||   :::
Db   3631 AGGCTTGAGTTGGCTCAGAAACTTAATGAAAATTATGAGGAAGTGAAATCTATAACCAAA 3690

Qy   1013 GlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsn 1032
          :::   |||   :::||||||               |||:::
Db   3691 GAAAGAAAAGTTCTAAAGGAA---------------TTACAGAAGTCATTTGAAACAGAG 3735

Qy   1033 GluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAsp 1052
             ::::::      :::||||||::::::                   |||||||||:::
Db   3736 AGAGACCACCTTAGAGGATATATAAGAGAAATTGAAGCTACAGGCCTACAAACCAAAGAA 3795

Qy   1053 AlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAla 1072
                                   :::   |||         |||:::   ::::::
Db   3796 GAACTAAAAATTGCTCATATTCACCTAAAAGAACACCAAGAAACTATTGATGAACTAAGA 3855

Qy   1073 MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProVal 1092
             ::::::      |||:::   |||
Db   3856 AGAAGCGTA------TCTGAGAAGACA--------------------------------- 3876

Qy   1093 LeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeu 1112
                |||:::               |||         |||:::||||||      |||
Db   3877 ------GCTCAAATAATAAATACTCAGGACTTAGAAAAATCCCATACCAAA------TTA 3924

Qy   1113 ProProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGly 1132
                |||:::                                    ||||||
Db   3925 CAAGAAGAGATCCCA---------------------------------GTGCTTCAT 3948

Qy   1133 AlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSer------Ser 1150
                :::   |||   :::         |||   |||:::|||:::         :::
Db   3949 GAGGAACAAGAGTTACTGCCTAATGTGAAAAAAGTCAGTGAGACTCAGGAAACAATGAAT 4008

Qy   1151 ArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIle 1170
             :::|||   ::::::      |||:::::::   :::   ::::::      ||||||
Db   4009 GAACTGGAGTTATTAACAGAACAGTCCACAACCAAGGACTCAACAACACTGGCAAGAATA 4068

Qy   1171 LysGlyArgLeuAspSerSerGluMetAspHisSerGlu---AsnGluAspTyrThrMet 1189
                                 ||||||:::            ||||||   :::
Db   4069 --------------------GAAATGGAAAGGCTCAGGTTGAATGAAAAATTTCAAGAA 4107

Qy   1190 SerSerProLeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgSer 1209
          |||                  :::   :::||||||:::               ::::::
Db   4108 AGTCAGGAAGAGATAAAATCTCTAACCAAGGAAAGAGACAACCTTAAAACGATAAAAGAA 4167

Qy   1210 GluLeuGluLysProArgGlyArgLysLysThrProValThrGlu--------------- 1224
             ||||||            :::   |||         :::   |||
Db   4168 GCCCTTGAAGTTAAACATGACCAGCTGAAAGAACATATTAGAGAAACTTTGGCTAAAATC 4227

Qy   1225 ------------------GlnGluGluLysLeuGlyMet------------AspAspLeu 1234
                            ||||||:::   |||   |||               ::::::
Db   4228 CAGGAGTCTCAAAGCAAACAAGAACAGTCCTTAAATATGAAAGAAAAAGACAATGAAACT 4287

Qy   1235 ThrLysLeuValGlnGluGln---------LysProLys--------------------- 1244
          ||||||:::|||   |||            |||||||||
Db   4288 ACCAAAATCGTGAGTGAGATGGAGCAATTCAAACCCAAAGATTCAGCACTACTAAGGATA 4347

Qy   1245 ---------------GlySerGlnArgSerArgLysArgGlyHis--------------- 1254
                         |||   :::         :::   |||
Db   4348 GAAATAGAAATGCTCGGATTGTCCAAAAGACTTCAAGAAAGTCATGATGAAATGAAATCT 4407
```

```
Qy       1255 ThrAlaSerGluSerAspGluGlnGln-----------------TrpProGluGluLys 1268
                 |||   |||  |||:::    |||                     |||
Db       4408 GTAGCTAAGGAGAAAGATGACCTACAGAGGCTGCAAGAAGTTCTTCAATCTGAAAGTGAC 4467

Qy       1269 ArgLeuLysGluAspIle--------------------LeuGluAsnGluAsp------ 1279
              :::|||||||||:::|||                    ||||||  |||:::
Db       4468 CAGCTCAAAGAAAACATAAAAGAAATTGTAGCTAAACACCTGGAAACTGAAGAGGAACTT 4527

Qy       1280 -----------------------GluGlnAsnSerProProLysLysGlyLysArgGly 1291
                                     ||||||               :::   :::
Db       4528 AAAGTTGCTCATTGTTGCCTGAAAGAACAAGAGGAAACTATTAATGAGTTAAGAGTGAAT 4587

Qy       1292 ArgProProLysProLeuGlyGlyGlyThrProLysGluGluProThrMetLysThrSer 1311
                       |||                       |||     |||::::::
Db       4588 CTTTCAGAGAAG-----------------GAAACTGAAATATCAACCATTCAAAAGCAG 4629

Qy       1312 LysLysGlySerLysLysLysSerGlyProProAlaProGluGluGluGluGluGluGlu 1331
                 :::                |||                 |||     |||:::||||||
Db       4630 TTAGAAGCAATCAATGATAAATTACAGAACAAGATCCAAGAGATTTATGAGAAAGAGGAA 4689

Qy       1332 -----------ArgGlnSerGlyAsnThrGluGlnLys---SerLysSerLysGlnHis 1346
                         :::|||      ::::::|||   ::::::    ||||||
Db       4690 CAATTTAATATAAAACAAATTAGTGAGGTTCAGGAAAAAGTGAATGAACTGAAACAATTC 4749

Qy       1347 ArgValSerArgArgAlaGlnGlnArgAla---GluSerProGluSerSerAlaIleGlu 1365
              :::       |||:::|||:::     |||     :::|||   ||||||        :::|||
Db       4750 AAGGAGCATCGCAAAGCCAAGGATTCAGCACTACAAAGTATAGAAAGTAAGATGCTCGAG 4809

Qy       1366 SerThrGlnSerThrProGlnLysGlyArg---------------------------- 1375
                       |||   |||   :::|||   |||
Db       4810 TT-GACCAACAGACTTCAAGAAAGTCAAGAAGAAATACAAATTATGATTAAGGAAAAAGA 4868

Qy       1376 -----------------GlyArgProSerLysThrProSerProSerGlnProLys 1388
                                |||  ||||||               |||::::::   |||
Db       4869 GGAAATGAAAAGAGTACAGGAGGCCCTTCAGATAGAGAGAGACCAACTGAAAGAAAA 4925
```

RESULT 9
US-09-885-535-3
; Sequence 3, Application US/09885535
; Patent No. US20020104105A1
; GENERAL INFORMATION:
; APPLICANT: Myriad Genetics, Inc.
; APPLICANT: Heichman, Karen
; APPLICANT: Bartel, Paul L.
; TITLE OF INVENTION: Protein-Protein Interactions
; FILE REFERENCE: 2318-266-II
; CURRENT APPLICATION NUMBER: US/09/885,535
; CURRENT FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/213,245
; PRIOR FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 9274
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-885-535-3

Alignment Scores:

Pred. No.:	7.58e-10	Length:	9274
Score:	231.50	Matches:	333
Percent Similarity:	32.97%	Conservative:	266
Best Local Similarity:	18.33%	Mismatches:	600
Query Match:	3.22%	Indels:	619
DB:	10	Gaps:	79

US-09-512-581B-2 (1-1391) x US-09-885-535-3 (1-9274)

```
Qy         22 IleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValValLysThr 41
                 :::   ::::::        |||    |||:::   |||:::|||:::   ::::::|||:::
Db        502 GTTAAAGAAGAGTACTTGAAAGCAGAACTTAGTAGGCAACTAGAAGGCATCTTAAAATCA 561

Qy         42 PheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAlaLeu 61
                 |||   |||::::::|||   :::|||                       |||
Db        562 TTTAAGGATGTTGAACAGAAAGCAGAG-------------------AATCATGTCCAG 600

Qy         62 HisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuValAla 81
              ||||||   |||                    |||
Db        601 CACCTTCAGTCGGCCTGTGCAAGCTCTCATCAA------------------------ 633

Qy         82 CysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAsp 101

Db        633 ---------------------------------------------------------- 633

Qy        102 LysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLys 121
                            |||   ::::::|||      :::          ||||||||
Db        634 --------------TTTCAGCAAATGTCTAGAGATTTTCAGGCTTGGCTGGATACAAAG 678

Qy        122 SerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTyr 141
                 :::   |||                                   ||||||:::
Db        679 AAAGAAGAGCAAAAC----------------------------------AAATCTCAT 702

Qy        142 AsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeu 161
                 |||      :::|||:::                                  |||
Db        703 CCAATATCTGCCAAACTCGAT-----------------------------GTCTTG 729

Qy        162 PheSerValIleAsnAsnGlyHisAsn-----------------GlnLysValHisMet 175
                 |||:::|||   :::                        :::   ||||||
Db        730 GAGTCATTAATTAAAGATCATAAAGACTTTAGTAAAACTTTGACCGCTCAGTCTCATATG 789

Qy        176 HisMetValAspLeuMetSerSerIleIleCysGluGlyAspThrVal----------- 191
              :::                  |||   ||||||:::   :::
Db        790 TAT--------------GAAAAAACCATTGCAGAAGGTGAAAATCTGTTATTAAAAACA 834

Qy        192 --------------------SerGlnGluLeuLeuAspThrValLeuValAsnLeuVal 204
                                    |||      |||:::|||:::      |||
Db        835 CAAGGGTCTGAGAAGGCAGCCTTACAGTTACAGCTTAATACAATTAAAACCAATTGGGAT 894

Qy        205 ProAlaHisLysAsnLeu-----------AsnLysGlnAlaTyrAspLeuAlaLysAla 220
                 :::|||   :::              ||||||         |||   ||||||
Db        895 ACATTTAATAAGCAGGTGAAAGAAAGAGAAAACAAGTTAAAAGAGTCATTGGAAAAAGCC 954

Qy        221 Leu---LeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGln 239
              |||      |||      :::   :::   |||:::|||                  |||
Db        955 CTTAAGTATAAAGAGCAAGTAGAGACTCTCTGGCCATGGATAGACAAATGCCAAAACAAC 1014

Qy        240 ValLeuMetLeu----------------------GlyLysThrSerIleSer------ 249
              :::      :::                      |||:::   ||||||:::
Db       1015 CTGGAGGAAATAAAATTTTGCTTGGATCCTGCTGAAGGAGAGAATTCTATTGCCAAGTTA 1074

Qy        250 --------------AspLeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsn 264
                            ::::::   :::|||   |||   ::::::   ||||||   |||
Db       1075 AAGTCTCTGCAGAAGGAAATGGACCAACAC---TTTGGTATGGTA---GAATTACTGAAC 1128

Qy        265 IleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerAsn 284
                 ::::::   ||||||||||||         :::::::::   ::::::   ::::::
Db       1129 AACACAGCCAATAGCTTGCTCAGTGTCTGT---GAGATAGATAAAGAAGTTGTTACAGAT 1185

Qy        285 AspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAsp 304
              :::||||||:::      :::|||   |||   ::::::   :::         :::|||
Db       1186 GAGAATAAGTCACTGATCCAGAAGGTGGACATGGTCACTGAACAACTTCACAGTAAGAAA 1245

Qy        305 ---------------------------------------SerGluLeuAlaSerGlnAsn 311
                                                     |||::::::   ::::::
Db       1246 TTCTGTCTGGAGAACATGACTCAGAAGTTTAAAGAATTTCAAGAAGTTTCCAAAGAATCT 1305

Qy        312 LysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIle------ 329
              |||      ||||||         :::      |||||||||         :::
Db       1306 AAAAGGCAGCTTCAGTGTGCAAAGGAGCAGCTA---GATATCCATGATTCGCTGGGATCC 1362

Qy        330 -------------------------ArgLeuGluCysValLysPheAlaSerHisCys 340
                                       |||:::         :::   :::
Db       1363 CAGGCTTACAGTAACAAATACCTGACCATGTTGCAAACTCAGCAGAAATCACTTCAGGCC 1422

Qy        341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
              |||   :::         ||||||||||||   |||   :::   |||   |||   :::
```

```
Db       1423 TTGAAGCATCAGGTAGATTTGGCTAAAAGACTTGCACAGGACCTTGTGGTAGAGGCCTCA 1482

Qy        361 AspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys 380
              |||   :::             ||||||::::::   :::   |||   |||::::::
Db       1483 GACTCAAAGGGAACC------TCTGATGTTTTATTACAAGTGGAAACCATAGCTCAAGAG 1536

Qy        381 AspIleLeuLeuValAsnAsp--------------------------------------- 387
                       |||
Db       1537 CATAGTACACTAAGTCAGCAGGTTGATGAAAAGTGTTCTTTCTTAGAAACCAAGCTTCAG 1596

Qy        388 ---------HisLeuLeuAsnPheValArgGlu-------------------------- 395
                       |||      |||   :::||||||
Db       1597 GGCATTGGGCATTTCCAGAATACCATTCGAGAAATGTTTTCTCAGTTCGCAGAGTTTGAT 1656

Qy        396 ------------------------------------ArgThrLeuAspLysArgTrp 402
                                                  ||||||   |||:::
Db       1657 GATGAACTGGATAGCATGGCTCCAGTGGGGAGAGATGCAGAAACATTGCAAAAGCAAAAG 1716

Qy        403 ArgValArgLys-----------------GluAlaMetMetGlyLeuAlaGlnIleTyr 416
                       |||                 ||||||:::|||
Db       1717 GAAACTATAAAAGCCTTTCTAAAGAAACTAGAAGCCCTCATGGCAAGCAATGACAATGCC 1776

Qy        417 LysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIle 436
                 |||                 |||   ::::::   :::      :::      |||
Db       1777 AATAAAACCTGCAAGATGATGTTAGCCACAGAAGAAACCTCTCCTGACCTTGTTGGAATC 1836

Qy        437 LysAsp------------------------------------------------------ 438
              |||
Db       1837 AAAAGGGACTTGGAGGCCTTAAGCAAACAATGCAACAAGTTACTGGACCGAGCCCAAGCC 1896

Qy        439 --------------------------LysLeuLeuHisIleTyrTyrGlnAsnSerIle 449
                                        |||   |||      :::|||
Db       1897 AGAGAAGAGCAGGTTGAAGGGACAATTAAGCGCCTTGAAGAATTTTACAGCAAATTGAAA 1956

Qy        450 AspAspArgLeuLeuValGluArgIle-----------PheAlaGlnTyrMetValPro 465
              :::      :::|||:::::::::                    :::|||      |||
Db       1957 GAATTTTCTATTCTGCTCCAGAAAGCCGAAGAACATGAAGAGTCACAAGGTCCTGTTGGT 2016

Qy        466 HisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAsp 485
                       ||||||      :::         ::::::   ::::::         :::
Db       2017 ATGGAAACGGAGACAATTAATCAGCAGCTTAACATGTTCAAGGTATTCCAGAAAGAAGAG 2076

Qy        486 Leu--------------------------------------------------------- 486
              :::
Db       2077 ATTGAACCCTTGCAAGGTAAACAGCAAGATGTAAACTGGTTAGGTCAAGGCCTTATTCAG 2136

Qy        487 AsnAlaValLysAla----------------------------------LeuAsnGlu 494
              :::|||   |||:::                                  :::|||
Db       2137 AGTGCTGCCAAAAGCACTAGCACTCAGGGCTTGGAGCATGACCTGGATGATGTCAATGCA 2196

Qy        495 MetTrpLys----------------------------------CysGlnAsnLeuLeu 502
                 ||||||                                     |||      |||
Db       2197 CGGTGGAAGACTCTCAATAAGAAGGTGGCTCAGCGAGCAGCCCAGCTGCAGGAGGCCTTG 2256

Qy        503 ArgHis------GlnValLysAspLeuLeuAspLeuIle-------------------- 513
                 |||      :::   :::|||   |||:::   :::
Db       2257 CTGCACTGTGGGAGGTTCCAGGATGCCCTGGAGTCCCTGCTCAGCTGGATGGTGGACACT 2316

Qy        514 --------------------LysGlnProLysThrAspAlaSer---ValLysAlaIle 525
                                  |||   |||      :::      |||||||||
Db       2317 GAGGAGCTTGTGGCCAATCAGAAGCCCCCGTCGGCTGAGTTCAAAGTGGTAAAGGCCCAG 2376

Qy        526 PheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPhe 545
                 :::      ::::::   |||   |||   |||                  :::
Db       2377 ATACAAGAACAAAAGCTTCTCCAGAGATTGTTGGATGACCGAAAATCTACGGTGGAGGTA 2436

Qy        546 MetLysLys-----------------PheThrGlnValLeuGluAspAspGluLysIle 559
              :::|||:::                    |||         |||      ||||||
Db       2437 ATCAAACGAGAAGGAGAAAAAATTGCTACAACAGCAGAGCCCGCAGATAAAGTGAAGATT 2496

Qy        560 ArgLysGlnLeuGluValLeuValSer-----------------------ProThrCys 571
                 |||||||||   :::|||   |||
Db       2497 TTGAAACAGCTCAGTCTCTTGGATAGCAGATGGGAGGCATTGCTTAATAAAGCTGAAACA 2556

Qy        572 SerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAsnProLys 591
                       :::|||   ||||||   :::      :::   ::::::
Db       2557 AGGAATCGTCAGTTGGAAGGT---ATCTCGGTGGTAGCACAGCAATTTCATGAAACCTTA 2613

Qy        592 GlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaProVal 611
              :::|||   |||   :::|||         |||:::         |||         |||:::
Db       2614 GAACCACTGAACGAGTGGCTTACAACCATAGAAAAGAGGCTGGTGAATTGTGAACCCATA 2673

Qy        612 HisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGly 631
                             :::         |||   |||   :::|||   :::
Db       2674 GGAACCCAAGCATCTAAACTTGAGGAACAAATTGCACAG---CACAAAGTTCTGCAAGAG 2730

Qy        632 ThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGlu 651
                             :::   |||   |||||||||:::         |||||||||
Db       2731 GACATCTTACTCAGGAAACAAAATGTA------GATCAGGCTTTACTAAATGGTTTAGAA 2784

Qy        652 LeuLeuLys-----------------ValLeuSerPheThrHisProIleSerPheHis 665
              |||||||||                 ||||||                  :::
Db       2785 CTACTTAAACAAACCACAGGTGATGAAGTTTTAATAATTCAAGATAAATTGGAAGCCATT 2844

Qy        666 SerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAla 685
                 |||         :::      :::      |||   |||      |||
Db       2845 AAAGCAAGGTACAAAGAC------ATTACTAAACTGAGCACTGATGTGGCCAAGACTCTG 2898

Qy        686 GluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHis 705
              |||   |||||||||:::   :::         |||      ||||||
Db       2899 GAACAGGCGCTGCAGCTTGCAAGGCGGCTGCACTCCACACACGAAGAGCTGTGTACCTGG 2958

Qy        706 IleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLysGlyProProArgGln 725
              :::         :::   :::|||   ::::::::::::
Db       2959 CTGGACAAAGTGGAGGTGGAATTACTTTCATATGAAACTCAG----------------- 3000

Qy        726 AlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAla 745
                                                  :::      :::|||      |||
Db       3001 ------------------------------GTTCTGAAAGGAGAAGAAGCAAGTCAAGCA 3030

Qy        746 GlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIle-Th 765
              |||:::      |||         :::   |||   :::   |||:::  ||
Db       3031 CAAATG---AGACCAAAGGAACTGAAAAAGGAAGCTAAGAACAACAAAGCCTTACTGGAC 3087

Qy        765 rProLeuValThrIleGlyHisIle---------------------------------- 773
              |||||||:::         :::
Db       3088 TCCCTTAATGAAGTGAGCAGTGCTTTGCTGGAACTGGTACCATGGAGGGCAAGAGAAGGA 3147

Qy        774 ---------------------AlaLeuLeuAlaProAspGlnPheAlaAlaPro----- 784
                                   :::   |||   |||      |||   |||
Db       3148 CTTGAGAAAATGGTAGCTGAGGACAATGAGCGCTACCGATTAGTGAGCGACACCATCACT 3207

Qy        785 -------TrpLysSerTrpValAlaThrPhe--------------IleValLysAspLe 797
                     |||:::   :::      |||                 :::   |||   ||
Db       3208 CAGAAGGTGGAGGAGATCGATGCAGCCATTCTGCGATCACAGCAGTTTGACCAAGCAGCT 3267

Qy        797 uLeuMetAsnAspArgLeuProGly-----------LysLysThrThrLysLeuTrpVa 813
              |:::::::::         ||||||            ||||||         ||||||||
Db       3268 GATGCTGAGT------TATCCTGGATTACTGAAACAGAAAAAAAATTGATGTCTCTGGGT 3321

Qy        813 lProAspGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetVa 833
              |                  :::|||      ::::::|||         |||
Db       3322 GACATCAGGCTT-GAGCAAGACCAGACTTCTGCTCAGCTTCAAGTTCAAAAG-------- 3372

Qy        833 lArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLe 853
                                                         |||   |||:::   ::
Db       3373 ------------------------------------------ACATTCACCATGGAGAT 3389

Qy        853 uLeuThrThrIleLeuHisSerAsp------GlyAspLeuThrGluGlnGly---LysIl 870
              :|||         |||   |||         ||||||   :::   |||   |||||
Db       3390 TTTGAGA---------CACAAGGATATTATTGATGACCTTGTTAAATCTGGGCATAAAAT 3440

Qy        870 e--------------------SerLysProAspMetSerArg----------------- 877
              |                    |||      |||   :::
Db       3441 CATGACCGCATGCAGTGAAGAGGAAAAGCAATCAATGAAGAAAAAACTGGACAAGGTACT 3500
```

```
Qy        878 -----------------------------------------LeuArgLeuAlaAlaGl 883
                                                        |||:::|||
Db       3501 GAAGAACTATGATACCATCTGCCAGATTAATTCAGAAAGGTATCTGCAGCTGGAACGGGC 3560

Qy        883 ySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIle---------IleTh 900
               ::::::|||           |||      |||   |||:::          :::||
Db       3561 ACAGTCCCTGGTTAACCAATTCTGGGAA---ACATATGAAGAACTTTGGCCATGGCTGAC 3617

Qy        900 rLeuGluGln---------TyrGlnLeuCysAlaLeuAlaIle---------------- 911
              |    |||         ||||||  |||   |||:::
Db       3618 AGAAACACAATCAATCATCTCTCAGCTTCCCGCCCCAGCCCTTGAATATGAAACTCTAAG 3677

Qy        912 -------AsnAspGluCysTyrGlnValArgGlnValPheAlaGln------------- 924
                     :::|||      |||:::|||::::::   |||:::
Db       3678 GCAGCAGCAGGAAGAACATCGGCAACTGCGTGAGTTGATAGCTGAACACAAGCCTCATAT 3737

Qy        925 ----LysLeuHisLys---GlyLeuSerArgLeuArgLeu---ProLeuGlu-------- 938
                  |||::::::|||   |||          |||   |||   |||   |||
Db       3738 AGATAAGATGAACAAAACTGGGCCACAGTTACTGGAATTGAGCCCTGGGGAAGGCTTTTC 3797

Qy        939 ------------TyrMetAlaIleCysAlaLeuCysAla------LysAspProValLy 952
                          |||:::|||      |||   :::      |||:::   |||||
Db       3798 TATCCAAGAGAAGTATGTGGCAGCCGACACCCTTTACAGTCAAATTAAAGAAGATGTCAA 3857

Qy        952 sGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTy 972
              |:::|||                     |||   :::   |||:::         :::::
Db       3858 AAAGCGTGCTGTG-----------GCACTGGATGAAGCCATTTCTCAATCAACTCAGTT 3905

Qy        972 rLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValVa 992
              :      :::          :::   |||   |||       ::::::
Db       3906 CCATGACAAGATAGATCAGATCCTTGAGAGCCTGGAACGCATCGTGGAACGTCTGAGGCA 3965

Qy        992 lProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGln-------- 1009
               |||   :::|||          :::   :::   :::   :::|||   |||
Db       3966 GCCACCCTCTATC---------TCTGCAGAGGTTGAGAAGATCAAGGAACAGATCAGTGA 4016

Qy       1010 ------------------AspIleGluGlnLeuLysAspValLysGluCysLeuTrpPh 1023
                                |||:::|||:::|||:::   :::   |||   |||
Db       4017 AAATAAGAATGTGTCAGTAGACATGGAAAAGCTACAGCCGTTGTATGAAACTCTTAAACA 4076

Qy       1023 eValLeuGluIleLeuMetAlaLys---------------------------------- 1031
                 |||    ::::::|||:::
Db       4077 GAGGGGAGAGGAAATGATTGCTAGATCTGGGGGGACTGATAAAGACATATCTGCCAAAGC 4136

Qy       1032 -------AsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGl 1049
                     ::::::   :::          ||||||   :::   :::      :::   ::
Db       4137 TGTTCAGGATAAGCTTGACCAAATGGTTTTCATTTGGGAGAACATACACACACTGGTGGA 4196

Qy       1049 nThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCy 1069
              :    ::::::|||:::       |||       ::::::   ||||||   :::      ||
Db       4197 AGAGAGGGAAGCCAAACTACTGGATGTGATGGAGCTAGCAGAAAAGTTCTGG------TG 4250

Qy       1069 sAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLy 1089
              ||||          |||:::      |||   |||          :::
Db       4251 TGATCACATGTCATTGATAGTTACCATTAAAGATACTCAAGATTTCATCCGGGACCTGGA 4310

Qy       1089 sAspProValLeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLy 1109
              :||||||   :::                   |||      :::                 ::
Db       4311 AGATCCTGGAATTGATCCTTCAGTAGTAAAACAACAGCAAGAAGCAGCAGAGACCATAAG 4370

Qy       1109 s---------AsnTyrLeuProProGluMetLysSerPhePheThr------------- 1121
              :         :::   |||      |||:::
Db       4371 GGAAGAAATAGATGGACTACAGGAGGAGCTGGATATAGTTATTAACCTAGGTTCTGAACT 4430

Qy       1122 ------------------ProGlyLysProLysThrThrAsnValLeuGlyAlaValAs 1135
                                |||   ||||||                :::      :::||
Db       4431 CATTGCGGCATGTGGGGAGCCTGATAAACCCATTGTCAAGAAGAGTATAGATGAGTTAAA 4490

Qy       1135 nLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrVa 1155
              |          |||          |||          :::   :::      ::::::|||
Db       4491 TTCAGCATGGGATTCTCTAAATAAAGCTTGGAAAGACCGGATTGACAAACTTGAGGAGGC 4550

Qy       1155 lSerAsnAlaSer------------------------------------------- 1159
                  |||:::
Db       4551 AATGCAGGCTGCCGTTCAGTACCAGGATGGACTGCAGGCGGTATTTGACTGGGTAGATAT 4610

Qy       1160 ---------------SerSerSerAsnProSerSerProGlyArgIleLysGlyArgLe 1174
                             :::|||   :::|||                      :::
Db       4611 TGCAGGTGGTAAATTAGCTTCAATGTCTCCAATTGGAACAGATCTCGAAACTGTCAAGCA 4670

Qy       1174 uAspSerSerGluMetAspHisSerGluAsnGluAspTyrThrMet------------- 1189
                    |||:::          ::::::|||   |||
Db       4671 GCAGATTGAAGAGCTAAAGCAATTTAAGTCTGAGGCCTATCAACAGCAGATAGAAATGGA 4730

Qy       1190 ------------SerSerProLeuProGlyLysLysSerAspLysArgAspAspSerAs 1205
                          :::   |||      ||||||      :::         |||
Db       4731 AAGACTGAATCATCAAGCAGAGCTTTTGCTAAAGAAAGTAACAGAAGAGAGTGACAAACA 4790

Qy       1205 pLeuValArgSer---------GluLeuGluLysProArgGlyArgLysLysThrProVa 1222
                  |||:::             ||||||:::                  :::      ::
Db       4791 CACTGTTCAAGACCCATTAATGGAACTGAAATTGATATGGGATAGCCTGGAGGAGAGAAT 4850

Qy       1222 lThrGluGlnGluGluLysLeu---------------------------------Gl 1230
               :      ::::::   ||||||
Db       4851 CATCAACAGACAGCATAAACTGGAGGGTGCTCTATTAGCCTTGGGTCAGTTCCAACATGC 4910

Qy       1230 yMetAspAspLeuThrLys--------------------LeuValGlnGluGlnLysPr 1243
               :::|||:::|||                           |||:::   |||||||||||
Db       4911 CCTGGATGAGCTCCTGGCATGGCTGACACACACCGAGGGCTTGCTAAGTGAGCAGAAACC 4970

Qy       1243 oLysGlySer-------------------------------------------------- 1246
               |    |||
Db       4971 TGTTGGAGGAGACCCTAAAGCCATTGAAATTGAACTTGCCAAGCATCATGTGCTCCAAAA 5030

Qy       1247 ---------------------------GlnArgSerArgLysArgGlyHisThrAlaSe 1257
                                         :::         |||   |||:::
Db       5031 TGATGTATTAGCCCATCAGTCCACAGTGGAAGCCGTTAATAAAGCAGGAAATGATCTAAT 5090

Qy       1257 rGluSerAsp---------------------------------------------GluGl 1262
                ||||||                                                  ||
Db       5091 TGAATCAAGTGCAGGAGAAGAAGCAAGCAACCTTCAGAACAAGCTAGAGGTTTTAAATCA 5150

Qy       1262 nGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAs 1282
              |:::|||                    :::::::::||||||      :::::::::
Db       5151 ACGCTGG--------------------CAAAATGTTTTGGAAAAAACAGAACAAAGGAA 5189

Qy       1282 nSerPro-----------------ProLysLysGlyLysArgGlyArg---------- 1292
                                      ||||||      |||
Db       5190 GCAGCAGCTGGATGGTGCCTTGCGCCAGGCCAAAGGGTTCCATGGCGAAATTGAGGATTT 5249

Qy       1293 ------------------------------------ProProLysProLeuGlyGly-- 1299
                                                  ||||||||||||||
Db       5250 GCAGCAGTGGCTGACTGACACGGAGCGTCATCTGTTGGCATCTAAACCGCTGGGAGGTTT 5309

Qy       1300 ----GlyThrProLysGlu------------------------------------------ 1304
                  |||   ||||||
Db       5310 ACCGGAAACAGCCAAGGAGCAGCTTAATGTCCATATGGAAGTCTGTGCTGCCTTTGAAGC 5369

Qy       1305 ----GluProThrMetLysThr---SerLysLysGlySerLysLysLysSerGlyProPr 1322
                  |||   |||   |||:::   :::||||||   :::         :::      ||
Db       5370 TAAAGAAGAAACATATAAGAGTCTGATGCAGAAAGGCCAGCAGATGCTTGCAAGATGCCC 5429

Qy       1322 oAlaProGluGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLy 1342
              |           |||         ::::::            |||   ::::::|||   ::
Db       5430 A---AAATCTGCAGAGACAAATATTGACCAAGACATAAATAACTTGAAAGAAAAATGGGA 5486

Qy       1342 sSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSer 1358
              :|||   :::   :::::::::   |||       :::      |||:::
Db       5487 ATCGGTGGAAACCAAACTCAATGAAAGGAAAACTAAACTGGAAGAGGCT 5535
```

RESULT 10
US-10-037-270-130
; Sequence 130, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom

```
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Zhang, Jie
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Xue, Aidong J.
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  Tillinghast, John
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2B
;  CURRENT APPLICATION NUMBER: US/10/037,270
;  CURRENT FILING DATE:  2002-01-04
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1104
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 130
;   LENGTH: 8503
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (91)..(8082)
US-10-037-270-130

Alignment Scores:
Pred. No.:              7.36e-10      Length:         8503
Score:                  231.00        Matches:        285
Percent Similarity:     35.41%        Conservative:   267
Best Local Similarity:  18.28%        Mismatches:     541
Query Match:            3.21%         Indels:         467
DB:                     14            Gaps:           67

US-09-512-581B-2 (1-1391) x US-10-037-270-130 (1-8503)

Qy        17 ProGlyValLysGluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLys 36
              |||   |||   |||:::|||            :::   |||   :::::::::|||   :::
Db      1087 CCTTATGTTAATGAGGTATCA---------ACTGATGAAGCTCTCCTGAAAAGGTATAGA 1137

Qy        37 MetValValLysThrPheMetAspMetAspGlnAspSerGluGlu--------------- 51
                      |||      ||||||:::   :::      ||||||
Db      1138 ---------AAAGAAATAATGGATCTTAAAAAACAATTAGAGGAGGTTTCTTTAGAGACG 1188

Qy        52 ---------------GluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
                            ||||||:::            |||                  |||
Db      1189 CGGGCTCAGGCAATGGAAAAAGACCAATTGGCCCAACTTTTGGAAGAA------AAAGAT 1242

Qy        67 PhePheLeuLysHisProGlyLysAspVal------------ArgLeuLeuValAlaCys 82
                      |||         :::   :::                  |||:::||||||
Db      1243 TTGCTTCAGAAAGTACAGAATGAGAAAATTGAAAACTTAACACGGATGCTGGTG------ 1296

Qy        83 CysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLys 102
                                                           ||||||
Db      1297 --------------------------------------------ACCTCTTCTTCC--- 1308

Qy       103 LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLysSer 122
                                     :::|||   |||            ::::::   |||:::
Db      1309 --------------------CTCACGTTGCAA-----------CAGGAATTAAAGGCT 1335

Qy       123 ProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsn 142
                :::      |||      :::   |||      |||      :::|||:::   |||
Db      1336 AAAAGAAAACGAAGAGTTACTTGGTGCCTTGGCAAAATTAACAAAATGAAGAACTCAAAC 1395

Qy       143 IleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPhe 162
                                                        |||         |||
Db      1396 ------------------------------------------TATGCAGATCAATTT 1410

Qy       163 SerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSer 182
             ::::::      |||            |||   |||         ::::::|||:::
Db      1411 AATATACCAACAAAT---ATAACAACAAAAACACATAAGCTTTCTATAAATTTATTACGA 1467

Qy       183 SerIle---------IleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrVal 199
                |||         :::|||               |||            |||||||||:::
Db      1468 GAAATTGATGAATCTGTCTGTTCAGAGTCTGATGTTTTCAGTAACACTCTTGATACATTA 1527

Qy       200 Leu---ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAla 218
                   :::         ||||||   |||   ||||||::::::            :::
Db      1528 AGTGAGATAGAATGGAATCCAGCAACAAAGCTACTAAATCAGGAGAATATAGAA------ 1581

Qy       219 LysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsn 238

Db      1581 ------------------------------------------------------------ 1581

Qy       239 GlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeu 258
                                    :::   ::::::   |||            :::|||
Db      1582 --------------------AGTGAGTTGAACTCACTTCGTGCTGACTATGATAATCTG 1620

Qy       259 IleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGlu 278
             :::|||               |||         |||   :::         ::::::|||
Db      1621 GTATTA---------------GACTATGAACAACTACGAACAGAAAAAAAAAAAATGGAA 1665

Qy       279 PheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLys 298
                |||||||||      ::::::   :::      :::   ::::::         |||
Db      1666 TTGAAATTAAAAGAAAAGAATGATTTGGATGAATTTGAGGCTCTAGAAAGAAAAACTAAA 1725

Qy       299 MetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyr 318
                         ||||||   |||:::
Db      1726 ------------AAAGATCAAGAGATGCAA----------------------------- 1743

Qy       319 LeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSer 338

Db      1743 ------------------------------------------------------------ 1743

Qy       339 HisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArg 358
                   |||::::::                  |||:::|||   ::::::   :::|||
Db      1744 ------CTAATTCATGAAATTTCGAACTTAAAGAATTTAGTTAAGCATCGAGAAGTATAT 1797

Qy       359 SerHisAspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAla 378
             :::   |||   |||         :::                              ::::::
Db      1798 AATCAAGATCTTGAGAATGAACTC-----------------------------AGTTCA 1827

Qy       379 LysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeu 398
                |||   ::::::|||         :::      :::      ::::::|||
Db      1828 AAAGTAGAGCTGCTTAGAGAAAAGGAAGACCAGATTAAGAAGCTACAGGAATACATAGAC 1887

Qy       399 AspLysArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLys 418
             ::::::                  :::      |||   |||:::
Db      1888 TCTCAAAAG---------CTAGAAAATATAAAAATGGACTTGTCA-------------- 1923

Qy       419 TyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAsp 438
             |||:::|||:::|||                     ||||||      :::      :::
Db      1924 TACTCATTGGAAAGCATTGAAGACCCAAAACAAATGAAGCAGACTCTGTTTGATGCTGAA 1983

Qy       439 LysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIle 458
                :::                                          |||      :::
Db      1984 ACTGTA------------------------------------GCCCTTGATGCCAAGAGA 2007

Qy       459 PheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyr 478
             :::   ::::::         |||||||||      |||:::||||||   |||
Db      2008 GAATCAGCCTTTCTTAGAAGTGAAAATCTGGAGTTGAAGGAGAAAATGAAAGAACTTGCA 2067

Qy       479 TyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCys 498
                   |||   ::::::                        |||::::::
Db      2068 ACTACATACAAGCAAATGGAA--------------------AATGATATTCAGTTATAT 2106
```

```
Qy        499 GlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThr 518
              |||:::    |||      :::    |||        :::||||||   |||:::
Db       2107 CAAAGCCAATTGGAGGCAAAAAAGAAAATGCAAGTTGATCTGGAGAAAGAA--------- 2157

Qy        519 AspAlaSerValLysAlaIlePheSerLysValMetValIleThrArg----------- 534
                       :::::::::   |||:::::::::         :::|||
Db       2158 ---------TTACAATCTGCTTTTAATGAGATAACAAAACTCACCTCCCTTATAGATGGC 2208

Qy        535 -----------------------AsnLeuProAspProGlyLysAlaGlnAspPheMet 546
                                     ||||||      ||||||    |||
Db       2209 AAAGTTCCAAAAGATTTGCTCTGTAATTTGGAATTGGAAGGAAAGATTACTGATCTTCAG 2268

Qy        547 LysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeu 566
              |||:::     :::    :::|||::::::|||    :::|||:::::::::    :::|||
Db       2269 AAAGAACTAAATAAAGAAGTTGAAGAAAATGAAGCTTTGCGGGAAGAAGTC---ATTTTG 2325

Qy        567 ValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLys 586
              :::|||                  |||                |||     :::   |||:::
Db       2326 CTTTCA---------GAATTGAAATCTTTACCTTCTGAAGTAGAAAGGCTGAGGAAAGAG 2376

Qy        587 LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGlu 606
              :::              |||    :::     |||   :::|||              :::
Db       2377 ATA------------CAAGACAAATCTGAAGAGCTCCATATAATAACATCAGAAAAAGAT 2424

Qy        607 ArgIleAlaProValHisIleAspThrGluSer---IleSerAlaLeuIleLysGlnVal 625
              ::::::          :::      ||||||   :::        |||::::::::::::
Db       2425 AAATTGTTTTCTGAAGTAGTTCATAAGGAGAGTAGAGTTCAAGGTTTACTTGAAGAAATT 2484

Qy        626 AsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAla 645
                 |||:::   |||       |||       :::             ||||||||
Db       2485 GGGAAAACAAAAGATGACCTAGCAACTACACAGTCGAATTATAAAAGCACTGATCAAGAA 2544

Qy        646 Ile--------------------------------------------------------- 646

Db       2545 TTCCAAAATTTCAAAACCCTTCATATGGACTTTGAGCAAAAGTATAAGATGGTCCTTGAG 2604

Qy        647 ---------ArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSer 663
                       |||       |||::::::    :::
Db       2605 GAGAATGAGAGAATGAATCAGGAAATAGTTAATCTCTCTAAA------------------ 2646

Qy        664 PheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp--------- 680
                       |||:::   |||:::|||   |||      ||||||   :::
Db       2647 ------GAAGCCCAAAAATTTGATTCGAGTTTGGGTGCTTTGAAGACCGAGCTTTCTTAC 2700

Qy        681 ---------------AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThr 695
                             ||||||    |||          ::::::      |||
Db       2701 AAGACCCAAGAACTTCAGGAGAAAACACGTGAGGTTCAAGAAAGACTA------AATGAG 2754

Qy        696 GlySerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHis 715
                 :::::::::|||       :::      |||   |||      |||
Db       2755 ATGGAACAGCTGAAGGAACAATTAGAAAATAGAGATTCTCCGCTGCAAACTGTAGAAAGG 2814

Qy        716 HisLysSer--------------------------------------------------- 718
                 |||:::
Db       2815 GAGAAAACACTGATTACTGAGAAACTGCAGCAAACTTTAGAAGAAGTAAAAACTTTAACT 2874

Qy        719 ---LysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePhe 737
                 :::|||          :::|||   :::   ::::::
Db       2875 CAAGAAAAAGATGATCTAAAACAACTCCAAGAAAGCTTGCAAATT--------------- 2919

Qy        738 SerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro 757
                    ::::::        :::   |||   :::   ::::::   ::::::|||
Db       2920 ---GAGAGGGACCAACTCAAAAGTGATATTCACGATACTGTTAACATGAATATAGATACT 2976

Qy        758 SerAsn---------------LeuGluHisLeuIleThrProLeuValThrIleGlyHis 772
                                   ||||||   |||             ||||||
Db       2977 CAAGAACAATTACGAAATGCTCTTGAGTCTCTGAAACAACATCAAGAAACAATTAATACA 3036

Qy        773 IleAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPhe 792
              :::                                   ||||||   ::::::
Db       3037 CTA-----------------------------------AAATCGAAAATTTCTGAGGAA 3060

Qy        793 IleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrp 812
              :::   ::::::|||   |||
Db       3061 GTTTCCAGGAATTTGCATATG--------------------------------------- 3081

Qy        813 ValProAspGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMet 832
                       ||||||   :::   ||||||         :::   |||
Db       3082 ---------GAGGAAAATACAGGAGAAACTAAAGATGAATTTCAG-------------- 3117

Qy        833 ValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg 852
                 :::   ::::::|||:::                              :::|||   :::
Db       3118 ---CAAAAGATGGTTGGCATAGATAAAAAACAGGATTTGGAAGCTAAAAATACC---CAA 3171

Qy        853 LeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLys 872
                 ||||||      :::      :::      :::   :::|||   ||||||
Db       3172 ACACTAACTGCAGATGTTAAGGATAATGAGATAATTGAGCAACAAAGGAAGATATTTTCT 3231

Qy        873 ------ProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAla 890
                    :::   :::   |||:::                  |||   |||   :::
Db       3232 TTAATACAGGAGAAAAATGAACTCCAACAAATGTTAGAGAGTGTTATAGCAGAAAAGGAA 3291

Qy        891 Gln-----------------GluProCysTyrHisGluIleIleThrLeuGluGlnTyr 904
              |||                 |||                |||         |||:::
Db       3292 CAATTGAAGACTGACCTAAAGGAAAATATTGAAATGACCATTGAAAACCAGGAAGAATTA 3351

Qy        905 GlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGln 924
              :::|||                 :::   ||||||   :::   :::::::::   ||||||
Db       3352 AGACTT------------CTTGGGGATGAACTTAAAAAGCAACAAGAGATAGTTGCACAA 3399

Qy        925 LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAla 944
              :::   :::      :::   :::                |||               |||
Db       3400 GAAAAGAACCATGCCATAAAGAAA---------GAAGGAGAGCTTTCTAGGACCTGTGAC 3450

Qy        945 LeuCysAla------LysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeu 962
                 |||      ::::::   :::|||||||:::         ::::::
Db       3451 AGACTGGCAGAAGTTGAAGAAAAACTAAAGGAAAAGAGCCAGCAACTCCAAGAAAAACAG 3510

Qy        963 ValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGlu--- 981
                 :::   :::|||||||:::   |||         :::          ::::::|||
Db       3511 CAACAACTTCTTAATGTACAAGAAGAGATGAGTGAGATGCAGAAAAAGATTAATGAAATA 3570

Qy        982 --------------------LysLeuLeuSerLeuLeuProGluTyrValValProTyr 994
                                  |||      |||   |||:::::
Db       3571 GAGAATTTAAAGAATGAATTAAAGAACAAAGAATTGACATTGGAACATATGGAAACAGAG 3630

Qy        995 ThrIleHisLeuLeuAlaHis---AspProAspTyrValLysValGlnAspIle---Glu 1012
                 :::   |||            :::   :::|||   :::|||:::   |||   :::
Db       3631 AGGCTTGAGTTGGCTCAGAAACTTAATGAAAATTATGAGGAAGTGAAATCTATAACCAAA 3690

Qy       1013 GlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsn 1032
              :::   |||   :::||||||               |||:::
Db       3691 GAAAGAAAAGTTCTAAAGGAA---------------TTACAGAAGTCATTTGAAACAGAG 3735

Qy       1033 GluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAsp 1052
                 ::::::          :::||||||::::::              |||||||||:::
Db       3736 AGAGACCACCTTAGAGGATATATAAGAGAAATTGAAGCTACAGGCCTACAAACCAAAGAA 3795

Qy       1053 AlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAla 1072
                                   :::   |||      |||:::   ::::::
Db       3796 GAACTAAAAATTGCTCATATTCACCTAAAAGAACACCAAGAAACTATTGATGAACTAAGA 3855

Qy       1073 MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProVal 1092
                 ::::::      |||:::   |||
Db       3856 AGAAGCGTA------TCTGAGAAGACA------------------------------- 3876

Qy       1093 LeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeu 1112
                    |||:::               |||      |||:::||||||         |||
Db       3877 ------GCTCAAATAATAAATACTCAGGACTTAGAAAAATCCCATACCAAA------TTA 3924

Qy       1113 ProProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGly 1132
                 |||:::                                          ||||||
Db       3925 CAAGAAGAGATCCCA-----------------------------------GTGCTTCAT 3948

Qy       1133 AlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSer------Ser 1150
                       :::   |||   :::         |||   |||:::|||:::         :::
```

```
Db       3949 GAGGAACAAGAGTTACTGCCTAATGTGAAAAAAGTCAGTGAGACTCAGGAAACAATGAAT 4008

Qy       1151 ArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIle 1170
                 :::|||   ::::::      |||::::::   :::   ::::::      ||||||
Db       4009 GAACTGGAGTTATTAACAGAACAGTCCACAACCAAGGACTCAACAACACTGGCAAGAATA 4068

Qy       1171 LysGlyArgLeuAspSerSerGluMetAspHisSerGlu---AsnGluAspTyrThrMet 1189
                                    ||||||:::         ||||||   :::
Db       4069 ---------------------GAAATGGAAAGGCTCAGGTTGAATGAAAAATTTCAAGAA 4107

Qy       1190 SerSerProLeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgSer 1209
              |||                       :::   :::||||||:::         ::::::
Db       4108 AGTCAGGAAGAGATAAAATCTCTAACCAAGGAAAGAGACAACCTTAAAACGATAAAAGAA 4167

Qy       1210 GluLeuGluLysProArgGlyArgLysLysThrProValThrGlu--------------- 1224
                 ||||||          :::   |||       :::   |||
Db       4168 GCCCTTGAAGTTAAACATGACCAGCTGAAAGAACATATTAGAGAAACTTTGGCTAAAATC 4227

Qy       1225 ------------------GlnGluGluLysLeuGlyMet------------AspAspLeu 1234
                                ||||||:::   |||   |||            ::::::
Db       4228 CAGGAGTCTCAAAGCAAACAAGAACAGTCCTTAAATATGAAAGAAAAAGACAATGAAACT 4287

Qy       1235 ThrLysLeuValGlnGluGln---------LysProLys--------------------- 1244
              ||||||:::|||   |||            |||||||||
Db       4288 ACCAAAATCGTGAGTGAGATGGAGCAATTCAAACCCAAAGATTCAGCACTACTAAGGATA 4347

Qy       1245 ---------------GlySerGlnArgSerArgLysArgGlyHis--------------- 1254
                             |||      :::      :::      |||
Db       4348 GAAATAGAAATGCTCGGATTGTCCAAAAGACTTCAAGAAAGTCATGATGAAATGAAATCT 4407

Qy       1255 ThrAlaSerGluSerAspGluGlnGln------------------TrpProGluGluLys 1268
                 |||   |||   |||:::   |||                     |||
Db       4408 GTAGCTAAGGAGAAAGATGACCTACAGAGGCTGCAAGAAGTTCTTCAATCTGAAAGTGAC 4467

Qy       1269 ArgLeuLysGluAspIle---------------------LeuGluAsnGluAsp------ 1279
              :::|||||||||:::|||                     ||||||   |||:::
Db       4468 CAGCTCAAAGAAAACATAAAAGAAATTGTAGCTAAACACCTGGAAACTGAAGAGGAACTT 4527

Qy       1280 ------------------------GluGlnAsnSerProProLysLysGlyLysArgGly 1291
                                      ||||||         :::   :::
Db       4528 AAAGTTGCTCATTGTTGCCTGAAAGAACAAGAGGAAACTATTAATGAGTTAAGAGTGAAT 4587

Qy       1292 ArgProProLysProLeuGlyGlyGlyThrProLysGluGluProThrMetLysThrSer 1311
                       |||                     |||   |||::::::
Db       4588 CTTTCAGAGAAG------------------GAAACTGAAATATCAACCATTCAAAAGCAG 4629

Qy       1312 LysLysGlySerLysLysLysSerGlyProProAlaProGluGluGluGluGluGluGlu 1331
                 :::                     |||         |||:::||||||
Db       4630 TTAGAAGCAATCAATGATAAATTACAGAACAAGATCCAAGAGATTTATGAGAAAGAGGAA 4689

Qy       1332 ------------ArgGlnSerGlyAsnThrGluGlnLys---SerLysSerLysGlnHis 1346
                          :::|||         ::::::|||      ::::::   ||||||
Db       4690 CAACTTAATATAAAACAAATTAGTGAGGTTCAGGAAAAAGTGAATGAACTGAAACAATTC 4749

Qy       1347 ArgValSerArgArgAlaGlnGlnArgAla---GluSerProGluSerSerAlaIleGlu 1365
              :::      |||:::|||:::      |||   :::|||   ||||||      :::|||
Db       4750 AAGGAGCATCGCAAAGCCAAGGATTCAGCACTACAAAGTATAGAAAGTAAGATGCTCGAG 4809

Qy       1366 SerThrGlnSerThrProGlnLysGlyArg------------------------------ 1375
                    |||   |||   :::|||   |||
Db       4810 TT-GACCAACAGACTTCAAGAAAGTCAAGAAGAAATACAAATTATGATTAAGGAAAAAGA 4868

Qy       1376 ------------------GlyArgProSerLysThrProSerProSerGlnProLys 1388
                                |||   ||||||         |||::::::   |||
Db       4869 GGAAATGAAAAGAGTACAGGAGGCCCTTCAGATAGAGAGAGACCAACTGAAAGAAAA 4925
```

RESULT 11
US-09-893-519A-140
; Sequence 140, Application US/09893519A
; Publication No. US20030027243A1
; GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig
; APPLICANT: MOORE, Jeffrey
; APPLICANT: BUURMAN, Ed T.
; APPLICANT: BRADLEY, John
; APPLICANT: DESILVA, Thamara
; APPLICANT: HARRIS, Sandra
; APPLICANT: KOMARNITSKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MOORE, Daniel
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: HAQ, Tariq
; APPLICANT: ZHU, Shuhao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/1G548-US2
; CURRENT APPLICATION NUMBER: US/09/893,519A
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,164
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/224,457
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 140
; LENGTH: 7992
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Human Genbank/Z15005.1
; DATABASE ENTRY DATE: 1993-01-10
; RELEVANT RESIDUES: (1)..(7992)
US-09-893-519A-140

```
Alignment Scores:
Pred. No.:               1.02e-09      Length:          7992
Score:                   229.00        Matches:         285
Percent Similarity:      35.61%        Conservative:    268
Best Local Similarity:   18.35%        Mismatches:      546
Query Match:             3.18%         Indels:          455
DB:                      11            Gaps:            66

US-09-512-581B-2 (1-1391) x US-09-893-519A-140 (1-7992)

Qy         17 ProGlyValLysGluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLys 36
                 |||   |||   |||:::|||            :::   |||   :::::::::|||   :::
Db        997 CCTTATGTTAATGAGGTATCA---------ACTGATGAAGCTCTCCTGAAAAGGTATAGA 1047

Qy         37 MetValValLysThrPheMetAspMetAspGlnAspSerGluGlu--------------- 51
                       |||      ||||||:::   :::      ||||||
Db       1048 ---------AAAGAAATAATGGATCTTAAAAAACAATTAGAGGAGGTTTCTTTAGAGACG 1098

Qy         52 ---------------GluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
                             ||||||:::                  |||               |||
Db       1099 CGGGCTCAGGCAATGGAAAAAGACCAATTGGCCCAACTTTTGGAAGAA------AAAGAT 1152

Qy         67 PhePheLeuLysHisProGlyLysAspVal------------ArgLeuLeuValAlaCys 82
                       |||         :::   :::                  |||:::||||||
Db       1153 TTGCTTCAGAAAGTACAGAATGAGAAAATTGAAAACTTAACACGGATGCTGGTG------ 1206

Qy         83 CysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLys 102
                                                         ||||||
Db       1207 ---------------------------------------------ACCTCTTCTTCC--- 1218

Qy        103 LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLysSer 122
                                   :::|||   |||                  ::::::   |||:::
Db       1219 ---------------------CTCACGTTGCAA------------CAGGAATTAAAGGCT 1245

Qy        123 ProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsn 142
                 :::      |||      :::   |||      |||      :::|||:::   |||
Db       1246 AAAAGAAAACGAAGAGTTACTTGGTGCCTTGGCAAAATTAACAAAATGAAGAACTCAAAC 1305

Qy        143 IleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPhe 162
```

```
                                                         |||     |||
Db       1306 ------------------------------------------TATGCAGATCAATTT 1320

Qy        163 SerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSer 182
              ::::::      |||          |||   |||         ::::::|||:::
Db       1321 AATATACCAACAAAT---ATAACAACAAAAACACATAAGCTTTCTATAAATTTATTACGA 1377

Qy        183 SerIle---------IleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrVal 199
                 |||          :::|||                 |||||||||:::
Db       1378 GAAATTGATGAATCTGTCTGTTCAGAGTCTGATGTTTTCAGTAACACTCTTGATACATTA 1437

Qy        200 Leu---ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAla 218
                    :::         ||||||  |||   ||||||::::::     :::
Db       1438 AGTGAGATAGAATGGAATCCAGCAACAAAGCTACTAAATCAGGAGAATATAGAA------ 1491

Qy        219 LysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsn 238

Db       1491 ------------------------------------------------------------ 1491

Qy        239 GlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeu 258
                                      :::  ::::::  |||                :::|||
Db       1492 --------------------AGTGAGTTGAACTCACTTCGTGCTGACTATGATAATCTG 1530

Qy        259 IleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGlu 278
              :::|||                  |||         |||   :::      ::::::|||
Db       1531 GTATTA---------------GACTATGAACAACTACGAACAGAAAAAGAAGAAATGGAA 1575

Qy        279 PheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLys 298
                 |||||||||         ::::::   :::      :::   ::::::        |||
Db       1576 TTGAAATTAAAAGAAAAGAATGATTTGGATGAATTTGAGGCTCTAGAAAGAAAAACTAAA 1635

Qy        299 MetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyr 318
                          ||||||  |||:::
Db       1636 -----------AAAGATCAAGAGATGCAA------------------------------ 1653

Qy        319 LeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSer 338

Db       1653 ------------------------------------------------------------ 1653

Qy        339 HisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArg 358
                    |||::::::              |||:::|||   ::::::   :::|||
Db       1654 ------CTAATTCATGAAATTTCGAACTTAAAGAATTTAGTTAAGCATCGAGAAGTATAT 1707

Qy        359 SerHisAspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAla 378
              :::   |||   |||      :::                              ::::::
Db       1708 AATCAAGATCTTGAGAATGAACTC----------------------------AGTTCA 1737

Qy        379 LysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeu 398
              |||      ::::::|||         :::       :::      ::::::|||
Db       1738 AAAGTAGAGCTGCTTAGAGAAAAGGAAGACCAGATTAAGAAGCTACAGGAATACATAGAC 1797

Qy        399 AspLysArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLys 418
                 ::::::               :::         |||   |||:::
Db       1798 TCTCAAAAG---------CTAGAAAATATAAAAATGGACTTGTCA-------------- 1833

Qy        419 TyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAsp 438
              |||:::|||:::|||                        ||||||   :::         :::
Db       1834 TACTCATTGGAAAGCATTGAAGACCCAAAACAAATGAAGCAGACTCTGTTTGATGCTGAA 1893

Qy        439 LysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIle 458
                 :::                                     |||      :::
Db       1894 ACTGTA---------------------------------GCCCTTGATGCCAAGAGA 1917

Qy        459 PheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyr 478
                 :::   ::::::               |||||||||   |||:::||||||   |||
Db       1918 GAATCAGCCTTTCTTAGAAGTGAAAATCTGGAGTTGAAGGAGAAAATGAAAGAACTTGCA 1977

Qy        479 TyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCys 498
                    |||      ::::::                     |||::::::
Db       1978 ACTACATACAAGCAAATGGAA--------------------AATGATATTCAGTTATAT 2016

Qy        499 GlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThr 518
              |||:::   |||      :::   |||      :::||||||   |||:::
Db       2017 CAAAGCCAATTGGAGGCAAAAAAGAAAATGCAAGTTGATCTGGAGAAAGAA--------- 2067

Qy        519 AspAlaSerValLysAlaIlePheSerLysValMetValIleThrArg------------ 534
                    :::::::::   |||:::::::::         :::|||
Db       2068 ---------TTACAATCTGCTTTTAATGAGATAACAAAACTCACCTCCCTTATAGATGGC 2118

Qy        535 ------------------------AsnLeuProAspProGlyLysAlaGlnAspPheMet 546
                                      ||||||         ||||||      |||
Db       2119 AAAGTTCCAAAAGATTTGCTCTGTAATTTGGAATTGGAAGGAAAGATTACTGATCTTCAG 2178

Qy        547 LysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeu 566
              |||:::         :::      :::|||:::::|||   :::|||:::::::::   :::|||
Db       2179 AAAGAACTAAATAAAGAAGTTGAAGAAAATGAAGCTTTGCGGGAAGAAGTC---ATTTTG 2235

Qy        567 ValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLys 586
              :::|||                  |||               |||      :::   |||:::
Db       2236 CTTTCA--------GAATTGAAATCTTTACCTTCTGAAGTAGAAAGGCTGAGGAAAGAG 2286

Qy        587 LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGlu 606
              :::            |||      :::      |||   :::|||               :::
Db       2287 ATA-----------CAAGACAAATCTGAAGAGCTCCATATAATAACATCAGAAAAAGAT 2334

Qy        607 ArgIleAlaProValHisIleAspThrGluSer---IleSerAlaLeuIleLysGlnVal 625
              ::::::               :::      ||||||   :::      |||::::::::::::
Db       2335 AAATTGTTTTCTGAAGTAGTTCATAAGGAGAGTAGAGTTCAAGGTTTACTTGAAGAAATT 2394

Qy        626 AsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAla 645
                 |||:::   |||      |||         :::               |||||||||
Db       2395 GGGAAAACAAAAGATGACCTAGCAACTACACAGTCGAATTATAAAGCACTGATCAAGAA 2454

Qy        646 Ile--------------------------------------------------------- 646

Db       2455 TTCCAAAATTTCAAAACCCTTCATATGGACTTTGAGCAAAAGTATAAGATGGTCCTTGAG 2514

Qy        647 ---------ArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSer 663
                       |||         |||::::::   :::
Db       2515 GAGAATGAGAGAATGAATCAGGAAATAGTTAATCTCTCTAAA------------------ 2556

Qy        664 PheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp--------- 680
                       |||:::   |||:::|||   |||      ||||||   :::
Db       2557 ------GAAGCCCAAAAATTTGATTCGAGTTTGGGTGCTTTGAAGACCGAGCTTTCTTAC 2610

Qy        681 --------------AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThr 695
                            ||||||               ::::::         |||
Db       2611 AAGACCCAAGAACTTCAGGAGAAAACACGTGAGGTTCAAGAAAGACTA------AATGAG 2664

Qy        696 GlySerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHis 715
                    :::::::::|||         :::      |||   |||      |||
Db       2665 ATGGAACAGCTGAAGGAACAATTAGAAAATAGAGATTCTCCGCTGCAAACTGTAGAAAGG 2724

Qy        716 HisLysSer--------------------------------------------------- 718
                 |||:::
Db       2725 GAGAAAACACTGATTACTGAGAAACTGCAGCAAACTTTAGAAGAAGTAAAAACTTTAACT 2784

Qy        719 ---LysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePhe 737
                 :::|||         :::|||   :::   ::::::
Db       2785 CAAGAAAAAGATGATCTAAAACAACTCCAAGAAAGCTTGCAAATT-------------- 2829

Qy        738 SerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro 757
                 ::::::         :::   |||   :::   ::::::   ::::::|||
Db       2830 ---GAGAGGGACCAACTCAAAAGTGATATTCACGATACTGTTAACATGAATATAGATACT 2886

Qy        758 SerAsn--------------LeuGluHisLeuIleThrProLeuValThrIleGlyHis 772
                                  ||||||   |||               ||||||
Db       2887 CAAGAACAATTACGAAATGCTCTTGAGTCTCTGAAACAACATCAAGAAACAATTAATACA 2946

Qy        773 IleAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPhe 792
              :::                                 ||||||   ::::::
Db       2947 CTA--------------------------------AAATCGAAAATTTCTGAGGAA 2970

Qy        793 IleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrp 812
              :::   ::::::|||   |||
Db       2971 GTTTCCAGGAATTTGCATATG--------------------------------------- 2991
```

```
Qy        813 ValProAspGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMet 832
                       ||||||   :::    ||||||        :::   |||
Db       2992 ---------GAGGAAAATACAGGAGAAACTAAAGATGAATTTCAG-------------- 3027

Qy        833 ValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg 852
                 :::   ::::::|||:::                          :::|||   :::
Db       3028 ---CAAAAGATGGTTGGCATAGATAAAAAACAGGATTTGGAAGCTAAAAATACC---CAA 3081

Qy        853 LeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLys 872
                 ||||||      :::      :::      :::   :::|||   ||||||
Db       3082 ACACTAACTGCAGATGTTAAGGATAATGAGATAATTGAGCAACAAAGGAAGATATTTTCT 3141

Qy        873 ------ProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAla 890
                    :::   :::   |||:::               |||   |||   :::
Db       3142 TTAATACAGGAGAAAAATGAACTCCAACAAATGTTAGAGAGTGTTATAGCAGAAAAGGAA 3201

Qy        891 Gln-----------------GluProCysTyrHisGluIleIleThrLeuGluGlnTyr 904
              |||                 |||            |||         |||:::
Db       3202 CAATTGAAGACTGACCTAAAGGAAAATATTGAAATGACCATTGAAAACCAGGAAGAATTA 3261

Qy        905 GlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGln 924
              :::|||                :::   ||||||      :::   :::::::::   ||||||
Db       3262 AGACTT-----------CTTGGGGATGAACTTAAAAAGCAACAAGAGATAGTTGCACAA 3309

Qy        925 LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAla 944
              :::   :::      :::   :::                   |||         |||
Db       3310 GAAAAGAACCATGCCATAAAGAAA---------GAAGGAGAGCTTTCTAGGACCTGTGAC 3360

Qy        945 LeuCysAla------LysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeu 962
                 |||         ::::::   :::|||||||:::            ::::::
Db       3361 AGACTGGCAGAAGTTGAAGAAAAACTAAAGGAAAAGAGCCAGCAACTCCAAGAAAAACAG 3420

Qy        963 ValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGlu--- 981
              :::      :::|||||||:::   |||         :::         :::::::|||
Db       3421 CAACAACTTCTTAATGTACAAGAAGAGATGAGTGAGATGCAGAAAAAGATTAATGAAATA 3480

Qy        982 --------------------LysLeuLeuSerLeuLeuProGluTyrValValProTyr 994
                                  |||         |||      |||::::::
Db       3481 GAGAATTTAAAGAATGAATTAAAGAACAAAGAATTGACATTGGAACATATGGAAACAGAG 3540

Qy        995 ThrIleHisLeuLeuAlaHis---AspProAspTyrValLysValGlnAspIle---Glu 1012
                 :::   |||               :::   :::|||   :::|||:::   |||   :::
Db       3541 AGGCTTGAGTTGGCTCAGAAACTTAATGAAAATTATGAGGAAGTGAAATCTATAACCAAA 3600

Qy       1013 GlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsn 1032
              :::   |||   :::||||||                  |||:::
Db       3601 GAAAGAAAAGTTCTAAAGGAA--------------TTACAGAAGTCATTTGAAACAGAG 3645

Qy       1033 GluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAsp 1052
              ::::::         :::||||||::::::               |||||||||:::
Db       3646 AGAGACCACCTTAGAGGATATATAAGAGAAATTGAAGCTACAGGCCTACAAACCAAAGAA 3705

Qy       1053 AlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAla 1072
                                         :::   |||      |||:::   ::::::
Db       3706 GAACTAAAAATTGCTCATATTCACCTAAAAGAACACCAAGAAACTATTGATGAACTAAGA 3765

Qy       1073 MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProVal 1092
                 ::::::         |||:::   |||
Db       3766 AGAAGCGTA------TCTGAGAAGACA------------------------------- 3786

Qy       1093 LeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeu 1112
                    |||:::                  |||         |||:::||||||      |||
Db       3787 ------GCTCAAATAATAAATACTCAGGACTTAGAAAAATCCCATACCAAA------TTA 3834

Qy       1113 ProProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGly 1132
                    |||:::                                          ||||||
Db       3835 CAAGAAGAGATCCCA---------------------------------GTGCTTCAT 3858

Qy       1133 AlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSer------Ser 1150
                 :::   |||   :::         |||   |||:::|||:::            :::
Db       3859 GAGGAACAAGAGTTACTGCCTAATGTGAAAAAAGTCAGTGAGACTCAGGAAACAATGAAT 3918

Qy       1151 ArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIle 1170
                 :::|||   ::::::         |||:::::::   :::   ::::::      ||||||
Db       3919 GAACTGGAGTTATTAACAGAACAGTCCACAACCAAGGACTCAACAACACTGGCAAGAATA 3978

Qy       1171 LysGlyArgLeuAspSerSerGluMetAspHisSerGlu---AsnGluAspTyrThrMet 1189
                                    ||||||:::               ||||||   :::
Db       3979 --------------------GAAATGGAAAGGCTCAGGTTGAATGAAAAATTTCAAGAA 4017

Qy       1190 SerSerProLeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgSer 1209
                 |||                     :::   :::||||||:::            ::::::
Db       4018 AGTCAGGAAGAGATAAAATCTCTAACCAAGGAAAGAGACAACCTTAAAACGATAAAAGAA 4077

Qy       1210 GluLeuGluLysProArgGlyArgLysLysThrProValThrGlu--------------- 1224
                 ||||||            :::   |||      :::   |||
Db       4078 GCCCTTGAAGTTAAACATGACCAGCTGAAAGAACATATTAGAGAAACTTTGGCTAAAATC 4137

Qy       1225 -----------------GlnGluGluLysLeuGlyMet-----------AspAspLeu 1234
                                 ||||||:::   |||   |||               ::::::
Db       4138 CAGGAGTCTCAAAGCAAACAAGAACAGTCCTTAAATATGAAAGAAAAAGACAATGAAACT 4197

Qy       1235 ThrLysLeuValGlnGluGln---------LysProLys------------------- 1244
              |||||||:::|||   |||            |||||||||
Db       4198 ACCAAAATCGTGAGTGAGATGGAGCAATTCAAACCCAAAGATTCAGCACTACTAAGGATA 4257

Qy       1245 --------------GlySerGlnArgSerArgLysArgGlyHis-------------- 1254
                            |||      :::      :::   |||
Db       4258 GAAATAGAAATGCTCGGATTGTCCAAAAGACTTCAAGAAAGTCATGATGAAATGAAATCT 4317

Qy       1255 ThrAlaSerGluSerAspGluGlnGln-----------------TrpProGluGluLys 1268
                 |||   |||   |||:::   |||                         |||
Db       4318 GTAGCTAAGGAGAAAGATGACCTACAGAGGCTGCAAGAAGTTCTTCAATCTGAAAGTGAC 4377

Qy       1269 ArgLeuLysGluAspIle--------------------LeuGluAsnGluAspGluGln 1281
              :::|||||||||:::|||                    ||||||   |||:::|||
Db       4378 CAGCTCAAAGAAAACATAAAAGAAATTGTAGCTAAACACCTGGAAACTGAAGAGGAACTT 4437

Qy       1282 AsnSerProProLysLysGlyLysArgGlyArgProPro---LysProLeuGlyGlyGly 1300
                                 |||                        |||
Db       4438 AAAGTTGCTCATTGTTGCCTGAAAGAACAAGAGGAAACTATTAATGAGTTAAGAGTGAAT 4497

Qy       1301 ThrProLysGluGluProThrMetLysThrSerLysLys--------------GlySer 1315
                    ::::::|||      :::   |||   :::|||
Db       4498 CTTTCAGAGAAGGAAACTGAAATATCAACCATTCAAAAGCAGTTAGAAGCAATCAATGAT 4557

Qy       1316 LysLysLysSerGlyProProAlaProGluGluGluGluGluGlu---------GluArg 1332
              |||   ::::::                  |||:::||||||:::            ::::::
Db       4558 AAATTACAGAACAAGATCCAAGAGATTTATGAGAAAGAGGAACAACTTAATATAAAACAA 4617

Qy       1333 GlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgValSerArgArgAla 1352
                 |||         |||      ::::::   ||||||      :::         |||:::|||
Db       4618 ATTAGTGAGGTTCAGGAAAACGTGAATGAACTGAAACAATTCAAGGAGCATCGCAAAGCC 4677

Qy       1353 GlnGlnArgAla---GluSerProGluSerSerAlaIleGluSerThrGlnSerThrPro 1371
              :::      |||      :::|||   ||||||         :::|||      |||   |||
Db       4678 AAGGATTCAGCACTACAAAGTATAGAAAGTAAGATGCTCGAGTT-GACCAACAGACTTCA 4736

Qy       1372 GlnLysGlyArg------------------------------------------------ 1375
              :::|||   |||
Db       4737 AGAAAGTCAAGAAGAAATACAAATTATGATTAAGGAAAAAGAGGAAATGAAAAGAGTACA 4796

Qy       1376 GlyArgProSerLysThrProSerProSerGlnProLys 1388
              |||   ||||||               |||::::::   |||
Db       4797 GGAGGCCCTTCAGATAGAGAGAGACCAACTGAAAGAAAA 4835
```

RESULT 12
US-10-133-013-260
; Sequence 260, Application US/10133013
; Publication No. US20030166903A1
; GENERAL INFORMATION:
; APPLICANT: Astromoff, Anna
; APPLICANT: Bandman, Olga
; APPLICANT: Cocks, Benjamin G.
; TITLE OF INVENTION: GENES ASSOCIATED WITH VASCULAR DISEASE

```
;  FILE REFERENCE: PA-0049 US
;  CURRENT APPLICATION NUMBER: US/10/133,013
;  CURRENT FILING DATE:  2002-04-25
;  PRIOR APPLICATION NUMBER: 60/287,067
;  PRIOR FILING DATE: 2001-04-27
;  NUMBER OF SEQ ID NOS: 271
;  SOFTWARE: PERL Program
; SEQ ID NO 260
;   LENGTH: 8491
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030166903A1 441283.5
US-10-133-013-260

Alignment Scores:
Pred. No.:              1.12e-09     Length:        8491
Score:                  229.00       Matches:       285
Percent Similarity:     35.61%       Conservative:  268
Best Local Similarity:  18.35%       Mismatches:    546
Query Match:            3.18%        Indels:        455
DB:                     12           Gaps:          66

US-09-512-581B-2 (1-1391) x US-10-133-013-260 (1-8491)

Qy         17 ProGlyValLysGluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLys 36
              |||  |||  |||:::|||          :::   |||   ::::::::|||   :::
Db       1087 CCTTATGTTAATGAGGTATCA---------ACTGATGAAGCTCTCCTGAAAAGGTATAGA 1137

Qy         37 MetValValLysThrPheMetAspMetAspGlnAspSerGluGlu-------------- 51
                       |||      ||||||:::   :::    ||||||
Db       1138 ---------AAAGAAATAATGGATCTTAAAAAACAATTAGAGGAGGTTTCTTTAGAGACG 1188

Qy         52 --------------GluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
                            ||||||:::          |||                |||
Db       1189 CGGGCTCAGGCAATGGAAAAAGACCAATTGGCCCAACTTTTGGAAGAA------AAAGAT 1242

Qy         67 PhePheLeuLysHisProGlyLysAspVal------------ArgLeuLeuValAlaCys 82
                       |||         :::   :::         |||:::||||||
Db       1243 TTGCTTCAGAAAGTACAGAATGAGAAAATTGAAAACTTAACACGGATGCTGGTG------ 1296

Qy         83 CysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLys 102
                                                          ||||||
Db       1297 --------------------------------------------ACCTCTTCTTCC--- 1308

Qy        103 LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLysSer 122
                                  :::|||   |||            ::::::   |||:::
Db       1309 --------------------CTCACGTTGCAA------------CAGGAATTAAAGGCT 1335

Qy        123 ProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsn 142
                 :::      |||      :::   |||      |||      :::|||:::   |||
Db       1336 AAAAGAAAACGAAGAGTTACTTGGTGCCTTGGCAAAATTAACAAAATGAAGAACTCAAAC 1395

Qy        143 IleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPhe 162
                                                          |||         |||
Db       1396 --------------------------------------------TATGCAGATCAATTT 1410

Qy        163 SerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSer 182
              ::::::              |||         |||         ::::::|||:::
Db       1411 AATATACCAACAAAT---ATAACAACAAAAACACATAAGCTTTCTATAAATTTATTACGA 1467

Qy        183 SerIle---------IleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrVal 199
                 |||         :::|||            |||         |||||||||:::
Db       1468 GAAATTGATGAATCTGTCTGTTCAGAGTCTGATGTTTTCAGTAACACTCTTGATACATTA 1527

Qy        200 Leu---ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAla 218
                    :::            ||||||   |||   ||||||::::::      :::
Db       1528 AGTGAGATAGAATGGAATCCAGCAACAAAGCTACTAAATCAGGAGAATATAGAA------ 1581

Qy        219 LysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsn 238

Db       1581 ------------------------------------------------------------ 1581

Qy        239 GlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeu 258
                                   :::   ::::::   |||               :::|||
Db       1582 --------------------AGTGAGTTGAACTCACTTCGTGCTGACTATGATAATCTG 1620

Qy        259 IleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGlu 278
              :::|||                  |||         |||   :::         ::::::|||
Db       1621 GTATTA---------------GACTATGAACAACTACGAACAGAAAAAGAAGAAATGGAA 1665

Qy        279 PheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLys 298
                 |||||||||         ::::::   :::         :::   ::::::      |||
Db       1666 TTGAAATTAAAAGAAAAGAATGATTTGGATGAATTTGAGGCTCTAGAAAGAAAAACTAAA 1725

Qy        299 MetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyr 318
                          ||||||   |||:::
Db       1726 ------------AAAGATCAAGAGATGCAA----------------------------- 1743

Qy        319 LeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSer 338

Db       1743 ------------------------------------------------------------ 1743

Qy        339 HisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArg 358
                    |||::::::                  |||:::|||   ::::::   :::|||
Db       1744 ------CTAATTCATGAAATTTCGAACTTAAAGAATTTAGTTAAGCATCGAGAAGTATAT 1797

Qy        359 SerHisAspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAla 378
              :::   |||   |||      :::                              ::::::
Db       1798 AATCAAGATCTTGAGAATGAACTC----------------------------AGTTCA 1827

Qy        379 LysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeu 398
              |||   ::::::|||         :::      :::         ::::::|||
Db       1828 AAAGTAGAGCTGCTTAGAGAAAAGGAAGACCAGATTAAGAAGCTACAGGAATACATAGAC 1887

Qy        399 AspLysArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLys 418
                 ::::::         :::         |||   |||:::
Db       1888 TCTCAAAAG---------CTAGAAAATATAAAAATGGACTTGTCA-------------- 1923

Qy        419 TyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAsp 438
              |||:::|||:::|||                  ||||||         :::      :::
Db       1924 TACTCATTGGAAAGCATTGAAGACCCAAAACAAATGAAGCAGACTCTGTTTGATGCTGAA 1983

Qy        439 LysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIle 458
                 :::                                       |||      :::
Db       1984 ACTGTA---------------------------------GCCCTTGATGCCAAGAGA 2007

Qy        459 PheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyr 478
                 :::   ::::::         |||||||||      |||:::||||||   |||
Db       2008 GAATCAGCCTTTCTTAGAAGTGAAAATCTGGAGTTGAAGGAGAAAATGAAAGAACTTGCA 2067

Qy        479 TyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCys 498
                    |||      ::::::                        |||::::::
Db       2068 ACTACATACAAGCAAATGGAA--------------------AATGATATTCAGTTATAT 2106

Qy        499 GlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThr 518
              |||:::   |||         :::   |||      :::||||||   |||:::
Db       2107 CAAAGCCAATTGGAGGCAAAAAAGAAAATGCAAGTTGATCTGGAGAAAGAA--------- 2157

Qy        519 AspAlaSerValLysAlaIlePheSerLysValMetValIleThrArg------------ 534
                       :::::::::   |||:::::::::         :::|||
Db       2158 ---------TTACAATCTGCTTTTAATGAGATAACAAAACTCACCTCCCTTATAGATGGC 2208

Qy        535 -----------------------AsnLeuProAspProGlyLysAlaGlnAspPheMet 546
                                     ||||||         ||||||      |||
Db       2209 AAAGTTCCAAAAGATTTGCTCTGTAATTTGGAATTGGAAGGAAAGATTACTGATCTTCAG 2268

Qy        547 LysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeu 566
              |||:::      :::   :::|||::::::|||   :::|||::::::::::   :::|||
Db       2269 AAAGAACTAAATAAAGAAGTTGAAGAAAATGAAGCTTTGCGGGAAGAAGTC---ATTTTG 2325

Qy        567 ValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLys 586
              :::|||                  |||               |||      :::   |||:::
Db       2326 CTTTCA---------GAATTGAAATCTTTACCTTCTGAAGTAGAAAGGCTGAGGAAAGAG 2376
```

```
Qy    587 LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGlu 606
          :::              |||      :::      |||    :::|||             :::
Db   2377 ATA------------CAAGACAAATCTGAAGAGCTCCATATAATAACATCAGAAAAAGAT 2424

Qy    607 ArgIleAlaProValHisIleAspThrGluSer---IleSerAlaLeuIleLysGlnVal 625
          ::::::             :::      ||||||    :::       |||::::::::::::
Db   2425 AAATTGTTTTCTGAAGTAGTTCATAAGGAGAGTAGAGTTCAAGGTTTACTTGAAGAAATT 2484

Qy    626 AsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAla 645
             |||:::    |||       |||        :::               ||||||||
Db   2485 GGGAAAACAAAAGATGACCTAGCAACTACACAGTCGAATTATAAAAGCACTGATCAAGAA 2544

Qy    646 Ile--------------------------------------------------------- 646

Db   2545 TTCCAAAATTTCAAAACCCTTCATATGGACTTTGAGCAAAAGTATAAGATGGTCCTTGAG 2604

Qy    647 ---------ArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSer 663
                   |||       |||::::::    :::
Db   2605 GAGAATGAGAGAATGAATCAGGAAATAGTTAATCTCTCTAAA------------------ 2646

Qy    664 PheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp--------- 680
                |||:::    |||:::|||    |||       ||||||      :::
Db   2647 ------GAAGCCCAAAAATTTGATTCGAGTTTGGGTGCTTTGAAGACCGAGCTTTCTTAC 2700

Qy    681 --------------AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThr 695
                        ||||||    |||          ::::::        |||
Db   2701 AAGACCCAAGAACTTCAGGAGAAAACACGTGAGGTTCAAGAAAGACTA------AATGAG 2754

Qy    696 GlySerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHis 715
                :::::::::|||             :::       |||    |||     |||
Db   2755 ATGGAACAGCTGAAGGAACAATTAGAAAATAGAGATTCTCCGCTGCAAACTGTAGAAAGG 2814

Qy    716 HisLysSer--------------------------------------------------- 718
             |||:::
Db   2815 GAGAAAACACTGATTACTGAGAAACTGCAGCAAACTTTAGAAGAAGTAAAAACTTTAACT 2874

Qy    719 ---LysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePhe 737
             :::|||             :::|||    :::    ::::::
Db   2875 CAAGAAAAAGATGATCTAAAACAACTCCAAGAAAGCTTGCAAATT--------------- 2919

Qy    738 SerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro 757
             ::::::             :::    |||    :::    ::::::    ::::::|||
Db   2920 ---GAGAGGGACCAACTCAAAAGTGATATTCACGATACTGTTAACATGAATATAGATACT 2976

Qy    758 SerAsn---------------LeuGluHisLeuIleThrProLeuValThrIleGlyHis 772
                               ||||||   |||              ||||||
Db   2977 CAAGAACAATTACGAAATGCTCTTGAGTCTCTGAAACAACATCAAGAAACAATTAATACA 3036

Qy    773 IleAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPhe 792
          :::                                   ||||||   ::::::
Db   3037 CTA-----------------------------------AAATCGAAAATTTCTGAGGAA 3060

Qy    793 IleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrp 812
          :::   :::::|||    |||
Db   3061 GTTTCCAGGAATTTGCATATG--------------------------------------- 3081

Qy    813 ValProAspGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMet 832
                   ||||||   :::    ||||||       :::    |||
Db   3082 ---------GAGGAAAATACAGGAGAAACTAAAGATGAATTTCAG--------------- 3117

Qy    833 ValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg 852
             :::   ::::::|||:::                      :::|||    :::
Db   3118 ---CAAAAGATGGTTGGCATAGATAAAAAAACAGGATTTGGAAGCTAAAAATACC---CAA 3171

Qy    853 LeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLys 872
             ||||||    :::       :::       :::    :::|||    ||||||
Db   3172 ACACTAACTGCAGATGTTAAGGATAATGAGATAATTGAGCAACAAAGGAAGATATTTTCT 3231

Qy    873 ------ProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAla 890
                :::    :::    |||:::              |||    |||     :::
Db   3232 TTAATACAGGAGAAAAATGAACTCCAACAAATGTTAGAGAGTGTTATAGCAGAAAAGGAA 3291

Qy    891 Gln-----------------GluProCysTyrHisGluIleIleThrLeuGluGlnTyr 904
                              |||             |||             |||       |||:::
Db   3292 CAATTGAAGACTGACCTAAAGGAAAATATTGAAATGACCATTGAAAACCAGGAAGAATTA 3351

Qy    905 GlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGln 924
          :::|||                :::    ||||||       :::    :::::::::    ||||||
Db   3352 AGACTT-------------CTTGGGGATGAACTTAAAAGCAACAAGAGATAGTTGCACAA 3399

Qy    925 LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAla 944
          :::    :::    :::    :::                  |||          |||
Db   3400 GAAAAGAACCATGCCATAAAGAAA---------GAAGGAGAGCTTTCTAGGACCTGTGAC 3450

Qy    945 LeuCysAla------LysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeu 962
             |||          :::::: · :::||||||:::                :::::
Db   3451 AGACTGGCAGAAGTTGAAGAAAAACTAAAGGAAAAGAGCCAGCAACTCCAAGAAAAACAG 3510

Qy    963 ValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGlu--- 981
             :::    ::::|||||||:::    |||          :::              ::::::|||
Db   3511 CAACAACTTCTTAATGTACAAGAAGAGATGAGTGAGATGCAGAAAAAGATTAATGAAATA 3570

Qy    982 --------------------LysLeuLeuSerLeuLeuProGluTyrValValProTyr 994
                              |||       |||     .|||:::::
Db   3571 GAGAATTTAAAGAATGAATTAAAGAACAAAGAATTGACATTGGAACATATGGAAACAGAG 3630

Qy    995 ThrIleHisLeuLeuAlaHis---AspProAspTyrValLysValGlnAspIle---Glu 1012
             :::    |||             :::    :::|||    :::|||:::    |||    :::
Db   3631 AGGCTTGAGTTGGCTCAGAAACTTAATGAAAATTATGAGGAAGTGAAATCTATAACCAAA 3690

Qy   1013 GlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsn 1032
          :::    |||    :::||||||                |||:::
Db   3691 GAAAGAAAAGTTCTAAAGGAA--------------TTACAGAAGTCATTTGAAACAGAG 3735

Qy   1033 GluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAsp 1052
             ::::::          :::||||||::::::                ||||||||:::
Db   3736 AGAGACCACCTTAGAGGATATATAAGAGAAATTGAAGCTACAGGCCTACAAACCAAAGAA 3795

Qy   1053 AlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAla 1072
                               :::    |||          |||:::    ::::::
Db   3796 GAACTAAAAATTGCTCATATTCACCTAAAAGAACACCAAGAAACTATGATGAACTAAGA 3855

Qy   1073 MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProVal 1092
          ::::::       |||:::    |||
Db   3856 AGAAGCGTA------TCTGAGAAGACA-------------------------------- 3876

Qy   1093 LeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeu 1112
             |||:::              |||          |||:::||||||    |||
Db   3877 ------GCTCAAATAATAAATACTCAGGACTTAGAAAAATCCCATACCAAA------TTA 3924

Qy   1113 ProProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGly 1132
             |||:::                                         ||||||
Db   3925 CAAGAAGAGATCCCA------------------------------------GTGCTTCAT 3948

Qy   1133 AlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSer------Ser 1150
                :::    |||    :::       |||    |||:::|||:::       :::
Db   3949 GAGGAACAAGAGTTACTGCCTAATGTGAAAAAAGTCAGTGAGACTCAGGAAACAATGAAT 4008

Qy   1151 ArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIle 1170
          :::|||    ::::::       |||::::::    :::    ::::::       ||||||
Db   4009 GAACTGGAGTTATTAACAGAACAGTCCACAACCAAGGACTCAACAACACTGGCAAGAATA 4068

Qy   1171 LysGlyArgLeuAspSerSerGluMetAspHisSerGlu---AsnGluAspTyrThrMet 1189
                                |||||||:::    ·    ||||||    :::
Db   4069 --------------------GAAATGGAAAGGCTCAGGTTGAATGAAAAATTTCAAGAA 4107

Qy   1190 SerSerProLeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgSer 1209
          |||                :::    :::||||||:::                ::::::
Db   4108 AGTCAGGAAGAGATAAAATCTCTAACCAAGGAAAGAGACAACCTTAAAACGATAAAAGAA 4167

Qy   1210 GluLeuGluLysProArgGlyArgLysLysThrProValThrGlu--------------- 1224
             ||||||             :::    |||       :::    |||
Db   4168 GCCCTTGAAGTTAAACATGACCAGCTGAAAGAACATATTAGAGAAACTTTGGCTAAAATC 4227

Qy   1225 -----------------GlnGluGluLysLeuGlyMet-----------AspAspLeu 1234
                           ||||||:::    |||    |||                ::::::
```

```
Db      4228 CAGGAGTCTCAAAGCAAACAAGAACAGTCCTTAAATATGAAAGAAAAAGACAATGAAACT 4287

Qy      1235 ThrLysLeuValGlnGluGln---------LysProLys-------------------- 1244
             ||||||:::|||   |||          ||||||||
Db      4288 ACCAAAATCGTGAGTGAGATGGAGCAATTCAAACCCAAAGATTCAGCACTACTAAGGATA 4347

Qy      1245 --------------GlySerGlnArgSerArgLysArgGlyHis--------------- 1254
                           |||       :::      :::     |||
Db      4348 GAAATAGAAATGCTCGGATTGTCCAAAAGACTTCAAGAAAGTCATGATGAAATGAAATCT 4407

Qy      1255 ThrAlaSerGluSerAspGluGlnGln-----------------TrpProGluGluLys 1268
                |||   |||   |||:::   |||                     |||
Db      4408 GTAGCTAAGGAGAAAGATGACCTACAGAGGCTGCAAGAAGTTCTTCAATCTGAAAGTGAC 4467

Qy      1269 ArgLeuLysGluAspIle--------------------LeuGluAsnGluAspGluGln 1281
             :::|||||||||:::|||                    ||||||   |||:::|||
Db      4468 CAGCTCAAAGAAAACATAAAAGAAATTGTAGCTAAACACCTGGAAACTGAAGAGGAACTT 4527

Qy      1282 AsnSerProProLysLysGlyLysArgGlyArgProPro---LysProLeuGlyGlyGly 1300
                                   |||                     |||
Db      4528 AAAGTTGCTCATTGTTGCCTGAAAGAACAAGAGGAAACTATTAATGAGTTAAGAGTGAAT 4587

Qy      1301 ThrProLysGluGluProThrMetLysThrSerLysLys--------------GlySer 1315
                   ::::::|||      :::   |||   :::|||
Db      4588 CTTTCAGAGAAGGAAACTGAAATATCAACCATTCAAAAGCAGTTAGAAGCAATCAATGAT 4647

Qy      1316 LysLysLysSerGlyProProAlaProGluGluGluGluGluGlu---------GluArg 1332
             |||   ::::::                  |||:::||||||:::          ::::::
Db      4648 AAATTACAGAACAAGATCCAAGAGATTTATGAGAAAGAGGAACAATTTAATATAAAACAA 4707

Qy      1333 GlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgValSerArgArgAla 1352
                |||         |||      ::::::   ||||||   :::        |||:::|||
Db      4708 ATTAGTGAGGTTCAGGAAAACGTGAATGAACTGAAACAATTCAAGGAGCATCGCAAAGCC 4767

Qy      1353 GlnGlnArgAla---GluSerProGluSerSerAlaIleGluSerThrGlnSerThrPro 1371
             :::      |||   :::|||   ||||||        :::|||      |||   |||
Db      4768 AAGGATTCAGCACTACAAAGTATAGAAAGTAAGATGCTCGAGTT-GACCAACAGACTTCA 4826

Qy      1372 GlnLysGlyArg--------------------------------------------- 1375
             :::|||   |||
Db      4827 AGAAAGTCAAGAAGAAATACAAATTATGATTAAGGAAAAAGAGGAAATGAAAAGAGTACA 4886

Qy      1376 GlyArgProSerLysThrProSerProSerGlnProLys 1388
              |||   ||||||                |||::::::   |||
Db      4887 GGAGGCCCTTCAGATAGAGAGAGACCAACTGAAAGAAAA 4925
```

RESULT 13
US-10-118-513A-11
; Sequence 11, Application US/10118513A
; Publication No. US20030039995A1
; GENERAL INFORMATION:
; APPLICANT: Taga, Tetsuya
; APPLICANT: Kimura, Naoki
; TITLE OF INVENTION: THE YS68 GENE INVOLVED IN PRIMITIVE HEMATOPOIESIS
; FILE REFERENCE: 06501-107US1
; CURRENT APPLICATION NUMBER: US/10/118,513A
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: PCT/JP00/05756
; PRIOR FILING DATE: 2000-08-25
; PRIOR APPLICATION NUMBER: JP 11-288738
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: JP 11-288739
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: JP 2000-123721
; PRIOR FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 7034
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (8)..(6736)
US-10-118-513A-11

```
Alignment Scores:
Pred. No.:                1.03e-09      Length:        7034
Score:                    228.00        Matches:       339
Percent Similarity:       32.29%        Conservative:  236
Best Local Similarity:    19.03%        Mismatches:    579
Query Match:              3.17%         Indels:        632
DB:                       14            Gaps:          78

US-09-512-581B-2 (1-1391) x US-10-118-513A-11 (1-7034)

Qy        50 GluGluGluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPhePheLeu 69
             |||   ||||||::::::            |||   |||   :::   :::   ||||||
Db      1098 GAGACAGAGAAGAAAGTATGAGAGAAGCTCTGTCTCCCGATACCAGCG------TTTCTG 1151

Qy        70 LysHisProGlyLysAspVal---------------ArgLeuLeuValAlaCysCysLeu 84
                |||||||||   :::                ||||||         |||
Db      1152 TCTTTACCTGGCAAGTGAATATATATGGACAGGGAAAGCCTTCTGTGTAT-TTAGGGCTA 1210

Qy        85 AlaAspIlePheArgIleTyrAlaProGluAlaPro------------------------ 96
                ||||||   |||   |||      :::   |||
Db      1211 TTTGACATAAATCGTTGGTATCATGCACAAATGCCCGATTCTTTAAGATCAGGAGAATCT 1270

Qy        97 -----------------------------------------------TyrThrSer 99
                                                               ||||||
Db      1271 CTGCATAATTGCTCTTATTTTGCGTTGTGGTCATTGGATTCGGTTGTAAGTAGGACTTCT 1330

Qy       100 ProAspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAsp 119
             |||      :::   ||||||   :::                     :::|||:::
Db      1331 CCACATCACATCTTGGACATACTAGTACATGAGAGGAGTTTAAACCGAGGGGTTCCTCCT 1390

Qy       120 ThrLysSerProGln---------PheAsn--ArgTyrPhe-----------TyrLeuL 132
             :::      |||            ||||||  :::||||||              :::|||
Db      1391 TCCTACCCACCTCCAGAGCAATTTTTAACCCAAGTACTTTTAATTTGATGCCACTTGT 1450

Qy       132 euGluAsnIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSerAsnG 152
                 :::   |||         |||:::|||:::         |||         :
Db      1451 TTGTTAGACTC---TGGAGTTATCCATGTAACTTGTGCTGGATTTCAGAAGGAGACTTTG 1507

Qy       152 luIlePheThrGlnLeuTyrArg-ThrLeuPheSerValIleAsnAsnGlyHisAsnGln 171
             ::||||||   ::::::         ||||||   ||||||   :::   :::|||:::
Db      1508 ACATTTTTAAAGAAATCAGGACCAACTCTTAATGAAGTCATTCCTGATAGTTATAATCGA 1567

Qy       172 LysValHisMetHisMetValAspLeuMetSerSerIleIleCysGluGlyAspThrVal 191
                                                                     :::
Db      1568 -----------------------------------------------------TGTCTT 1573

Qy       192 SerGlnGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAsn 211
                      ||||||         |||:::::::::   |||:::         :::
Db      1574 GTTGCTGGTCTCCTCTCACCAAGACTTATTGATATTCAGCCTTCCAGTTTAAGTCAAGAA 1633

Qy       212 LysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAla---IleGlu 230
             :::|||         :::      |||   ::::::   ::::::            :::
Db      1634 GAACAATTAGAAGCTATATTGTCAGCAGCAATTCAGACAAGTTCCTTGGGACTTTTGACT 1693

Qy       231 ProTyrIleThrThrPhe----------------------------------------- 236
                ||||||   |||:::
Db      1694 GGTTACATCAGAACATGGATAATAGAAGAACAACCAAATTCTGCTGCTAATCTACGATTT 1753

Qy       237 ---------------PheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAspLeu 251
                            :::|||:::|||   :::|||   |||               |||
Db      1754 GTTCTTGAGTGGACATGGAATAAAGTG---GTTCTCACAAAAGAAGAGTTTGATAGGCTT 1810

Qy       252 SerGluHisValPheAsp---------LeuIleLeuGluLeuTyrAsnIleAspSer--- 267
                :::||||||              ::::::         |||   |||
Db      1811 TGTGTGCCGCTGTTTGACGGTTCGTGTCGTTTTATTGACCCACAGACTATTCAGTCTATC 1870

Qy       268 ---------HisLeuLeuLeuSerValLeuProGlnLeuGlu------PheLysLeuLys 282
                      |||||||||||||||   |||      :::            |||   ::::::
Db      1871 CAGCAGTGCCATTTACTGCTTAGCAACCTTAGTACAGTCTTAAGCTGTTTTGCAATGGAG 1930
```

```
Qy        283 SerAsn-----------------------AspAsnGluGluArgLeuGlnValValLys 294
              :::                          |||       :::    |||   :::
Db       1931 GCCCAGGGTATCACTGAGAGAGGACTGGTGGACTTGAGCAACAAGCACATGGTCACCCAG 1990

Qy        295 LeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeu 314
              ||||||    :::                                          |||
Db       1991 CTTCTCTGTCAGTAC-----------------------------GCACACATGGTTCTG 2020

Qy        315 TrpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysVal 334
              |||   |||:::   |||                  :::|||    |||:::
Db       2021 TGGTTCTGCCACTCGGGGCTT---------------CTGCCCGAAGGCTTAGATGATGCT 2065

Qy        335 LysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyr 354
                       |||       |||:::|||   :::            |||
Db       2066 CTGCACCTGTCAAGACTACGCTACAACTACCCTGTAATTCAGAACTACTATACAAGTCGT 2125

Qy        355 LeuLysValArgSerHisAspProGluGluAlaIleArgHisAspValIleValSerIle 374
                 :::          |||          ||||||
Db       2126 CGGCAGAAGTCTGAGCGCTCACCCAGAGGGAAGTGGAACCACGAC--------------- 2170

Qy        375 ValThrAlaAlaLysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArg 394
                                   |||:::   |||   |||::::::   :::
Db       2171 -----------------------TGCTTGATGATTGATGGATTAGTCTCTCAACTAGGA 2206

Qy        395 GluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMetMetGly--------- 411
              :::   ::::::|||   |||:::   |||   |||         |||
Db       2207 GATGAA---GTTGAGAAGTTGTGGAAG---CGGGACGAAGGTGGCACAGGAAGATACCCT 2260

Qy        412 -----------------LeuAlaGlnIleTyr------------------------Lys 417
                               |||      ||||||
Db       2261 CCTGCTAGCATCCACGCATTACTTGATATATATTTATTAGACAACATTACCGAAGCAAGC 2320

Qy        418 LysTyrAla-------------------------LeuGlnSerAlaAlaGlyLys--- 427
              |||:::|||                          :::   |||         |||
Db       2321 AAACATGCTATTACCATTTATTTGCTGCTTGATATTATGTATTCCTTTCCAAATAAAACG 2380

Qy        428 ---------------------AspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu 440
                                   |||      |||:::|||   :::   ||||||
Db       2381 GATACCCCCATTGAATCTTTCCCCACTGCCTTTGCTATTTCTGGGGCCAAGTTAAGCTA 2440

Qy        441 Leu--------------------HisIleTyrTyrGlnAsnSerIleAspAspArgLeu 453
              :::                    |||      |||:::|||   :::|||      |||
Db       2441 GTTCAAGGATTTTGGCTACTAGATCATAATGACTATGAGAATGGTTTAGAC------CTT 2494

Qy        454 LeuValGluArgIlePheAlaGln--------------------TyrMetValProHis 466
              |||         :::   |||:::                    ::::::
Db       2495 CTGTTTCACCCAGTTACTGCAAAGCCTGCATCGTGGCAACATTCAAAGATAATTGAAGCT 2554

Qy        467 AsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyr--------------- 481
                 :::         |||      :::   |||   ||||||
Db       2555 TTTATGAGTCAGGGAGAGCACAAACAGGCTCTCCGGTATCTTCAGACAATGAAGCCAACA 2614

Qy        482 -----------------AlaThrLeuAspLeuAsnAlaVal---------LysAlaLeu 492
                                |||   |||         :::          :::   :::
Db       2615 GTGTCCAGTAGCAATGAAGTTATCCTTCACCTCACTGTTCTACTTTTAATAGATGCATG 2674

Qy        493 AsnGluMetTrpLysCys------GlnAsnLeuLeuArgHisGlnValLysAspLeuLeu 510
                 |||   |||            ||||||      |||      ::::::::::||||||
Db       2675 GTTGAGGCCTGGAACTTACTGCGACAGAATTCAAACAGAGTAAATATAGAGGAATTATTA 2734

Qy        511 ----------------------------------------AspLeuIleLysGlnPro 516
                                                      ||||||:::|||   |||
Db       2735 AAGCACGCTTATGAAGTTTGTCAGGAGATGGGCTTAATGGAGGATTTACTGAAGCTGCCA 2794

Qy        517 LysThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeu 536
               |||:::
Db       2795 TTTACAAACACTGAG--------------------------------------------- 2809

Qy        537 ProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAsp 556
                                |||:::   :::   ||||||          |||:::
Db       2810 -----------------CAGGAATGCTTAGTGAAATTT---------TTACAGTCCAGT 2842

Qy        557 GluLysIleArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAla 576
                 :::                         :::|||   ||||||        :::|||   |||
Db       2843 ACCAGTGTTGAGAATCATGAATTCCTTCTAGT-TCACCATTTACAGCGTGC---CAATTA 2898

Qy        577 GluGlyCysValArg--------------------------------------------- 581
                    |||:::
Db       2899 TATTTCTGCCTTGAAACTAAACCAGATTCTGAAGAATAATCTCATGAGTGATCGTGACCC 2958

Qy        582 GluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGlu-MetIl 601
                 |||      |||:::|||:::   |||               |||      |||  ::::
Db       2959 TCGATTGCGGGAAAGATCGGTGACTCGAAATTCTATATTAGACCAGTATGGGAAAATCCT 3018

Qy        601 e---------------LysPheLeuLeuGluArgIleAlaProValHisIleAspThrGl 616
              :                |||      :::||||||   |||   |||:::   |||
Db       3019 ACCTAGAGTCCAGAGAAAGTTAGCTGTTGAGCGAGCTAAGCCTTACCACCTGTCGACATC 3078

Qy        616 uSerIle-----------------------------SerAlaLeuIleLysGlnValAs 626
               |||:::                             ||||||      |||
Db       3079 CTCAGTTTTTCATGAAGTTTCTAGACCCAAACCGTTATCGGCATTTCCAAAG-------- 3130

Qy        626 nLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValPro--------------- 641
               |||:::|||   ||||||                  :::
Db       3131 -AAAGCTATAACTGGAACAGTGTTAACCCGATCTACGTTCATCAGCAATGTTTTATCTAA 3189

Qy        642 ----------------ThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLys----- 654
                                              |||   |||:::   |||
Db       3190 AATTGGAGAGGTGTGGGCAAGTCATGAGCCTAGAAATGGCGTCTCACTTTTAACAGTCC 3249

Qy        655 -------------------------ValLeuSerPheThrHisProIleSerPheHisSe 666
                                        |||   ||||||   ||||||
Db       3250 TAAAACAGAACAGCCATCTCCTGTAGTACACTCTTTCCCACACCCA-------------- 3295

Qy        666 rAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGl 686
                                                            ::::::   ||
Db       3296 -------------------------------------------------GAGCTTCCTGA 3306

Qy        686 uAlaAlaLeuGlnIle---PheLysAsnThrGlySerLysIleGluGluAspPheProHi 705
               ||||    :::                  ||||||      :::|||
Db       3307 GGCGTTTGTTGGAACTCCAATTTCAAATACATCCCAGAGAATT----------------- 3349

Qy        705 sIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLysGlyProProArgGl 725
                    |||   ||||||   ::::::   |||                  |||   ||
Db       3350 -------TCTAGATTACTGGATTTGGTTGTCCAT--------------CCTGTACCCCA 3387

Qy        725 nAlaLysTyrAlaIleHisCysIleHis-----------------------AlaIlePh 737
               |           :::      |||                              :::
Db       3388 GCCTTCTCAGTGTTTGGAGTTTATTCAACAAAGTCCCACAAGATCTCCTTTGTGTCTGCT 3447

Qy        737 eSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPr 757
               ||||||                 :::|||      |||         :::      ||
Db       3448 GTCCAGTTCGTTACCATTAAGTTCACAGTTTAAAAGGCCACATCAGAATACCTCCAGGCC 3507

Qy        757 oSerAsnLeuGluHisLeuIleThrProLeuValThr------IleGlyHisIleAlaLe 775
               ||||   |||      |||   |||||||||:::                :::|||||
Db       3508 TTCAGAGTTGCTTTTACTTGAGACTCCTCTCATAGTTAAGAAAGCTAAATCTTTGGCTCT 3567

Qy        775 uLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLy 795
               |   |||                              |||      |||   |||
Db       3568 GTCAGCCACG--------------------TCTTCTGGATTTGCCGAGTTTACTCCTCC 3606

Qy        795 sAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeu----------- 811
                 :::|||   :::                |||   :::||||||   |||
Db       3607 ATCCATCCTTAGGTCT------------GGTTTTCGAACAACACCTTTAGCATCTCCCTC 3654

Qy        812 -TrpValProAspGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMe 831
                 |||         ::::::|||            ||||||         |||
Db       3655 TTTGTCACCTGGAAGATCTCTCACTCCGCCTTTCAGAGTTAAAGAAACAAGGATTTCATT 3714

Qy        831 tMetValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThr-------- 848
               |||                  ||||||      :::   :::         |||
Db       3715 CATGGAAGAA---------GGCATGAATACACACTGGACTGATAGAGCTACAGATGACCG 3765

Qy        849 -SerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGly---------------- 862
                :::|||         ::::::|||      |||      |||
```

```
Db     3766 AAATACAAAAGCGTTTGTTAGCACATCTTTCCATAAATGTGGACTTCCAGCAGAAACTGA 3825

Qy      863 ----------------------------------------AspLeu-------------- 864
                                                    |||:::
Db     3826 GTGGATGAAGACCAGTGATAAGAATACATATTTCCTCTGGATGTCCCTGCAAAGGGCCC 3885

Qy      865 ----------------------------ThrGluGlnGlyLysIleSerLysProAspMe 875
                                        |||      |||::::::   |||   |||::
Db     3886 TCAGAAAGTGGTGGCAGAGTCACTGGCTACCCATTCAGGAAGGCTGGAGAAACTGGATGT 3945

Qy      875 tSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTy 895
            :|||:::            |||   :::   :::   :::  |||           ||
Db     3946 GAGCAAAGAAGACAGCACAGCTTCCACCAGGTCAGACCAGACCTCCTTAGAG------TA 3999

Qy      895 rHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIle----------- 911
             ||||:::      :::   |||      :::   |||         :::
Db     4000 TCATGACGCACCATCACCAGAAGACTTGGAAGG-TGCTGTTTTTGTGTCTCCCAAGCCAG 4058

Qy      912 ----------------------------------------AsnAspGluCysTyrGlnVa 918
                                                    |||   |||         ::
Db     4059 CATCTTCCTCCACTGAACTAACTACTAATTCAACTCTACAAACAGAGAGGG-ATAATGAT 4117

Qy      918 lArgGlnValPheAlaGlnLys------LeuHis-------------------------- 927
            ::::      :::||||||      ||||||
Db     4118 AAAGATGCGTTTAAGTCAGAAGGTACTCCTTCACCCGTGAAGAAACAAATAGGCACGGGA 4177

Qy      927 ------------------------------------------------------------ 927

Db     4178 GACGCTGCAGTGGAAGCATTTTCAGAACTGAGTCGCTTAGACCCTGTTGAAAGAGCTGAA 4237

Qy      928 ----------------------------LysGlyLeuSerArgLeuArgLeuProLeuGl 938
                                        ||||||         |||:::   |||
Db     4238 GCTTCTTTTGGTGTGTCGTCAGTCTGTGAAGGGGAAACCTCCACTTCAAACTCCAAGACG 4297

Qy      938 u------TyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArg----- 954
                   :::|||      |||   |||   |||:::   |||      ::::::
Db     4298 TCAGTTCTGGATGGAATCGTGCCTATTGAGAGCCGAACCTCCATACTTACAGCAGACCAC 4357

Qy      954 ------------------------------------------------------------ 954

Db     4358 AAAGAGTCTGTGGCCAACACGGTTGCAGATGTTGAAAGCTCTGGGTCCACCAGCTCCAAG 4417

Qy      955 ----ArgAlaHisAlaArgGlnCysLeuValLysAsn---------------------Il 966
                |||                        ||||||                     ::
Db     4418 TGCCCGGTTACCTCTGAACGCAGCCTCGGCCAAAAACTAACATTAAACTTAAAAGAAGAT 4477

Qy      966 eAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSerLe 986
            :      :::      |||   ::::::               :::|||   |||
Db     4478 GAAATAGAAGCTCATGTACCAAGGAGAACGTTGGTTTACCAGAAGAAAGCCCTCGAATT 4537

Qy      986 uLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrVa 1006
             |||         ::::::   :::   |||||||||:::            :::   ::
Db     4538 TCTGCTGCTCCTTCTGATACT-CACGAGATTCATCTAATTGGA------TGTGAAAATCT 4590

Qy     1006 lLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValLeuGl 1026
            ::::|||||||:::   |||:::      :::   |||            :::      |||
Db     4591 TGAAGTTCAAAATTCAGAAGAG------GAGGCCAAGAATCTTTCATTTGATGAGTTGTA 4644

Qy     1026 uIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAs 1046
             |||   |||:::                  :::   |||   :::          :::
Db     4645 TCCCTTAGGGGCAGAGAAACTTGAGTATAATCTCAGTACTATTGAGCAGCAGTTTTGTGA 4704

Qy     1046 nIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTy 1066
            ::::         ||||||:::         |||   |||:::::::::   :::|||::
Db     4705 CTTGCCTGATGACAAAGACTCTGCTGAATGTGATGCTGCTGAAGTAGACGGGGAACTTTT 4764

Qy     1066 rThrValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSerLeuGl 1086
            :                           :::         :::   |||      |||||
Db     4765 T---------------------------GTGGCCCAGAGCAACTTTACCCTGATTTTAGA 4797

Qy     1086 uSerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsnPheSe 1106
            |      :::                                 :::   :::   :::   ||
Db     4798 AGGTGAAGAAGGA---------------------------GAAGCTGAGGCAAGCGACTC 4830

Qy     1106 rAsnThrLysAsnTyrLeuProProGluMetLysSerPhePheThrProGlyLysPro-- 1125
            |         |||   ||||||               |||                 ||||||
Db     4831 TGCAGCACCTAATATGTTACCGAAATCGACCAAGGAA---------------AAACCTGT 4875

Qy     1126 ----------------------------LysThrThrAsnValLeuGlyAlaValAsnLy 1136
                                        :::   |||::::::         ||||||
Db     4876 GTGCTACAGGGAACCCCATAATCAGGAGCGCGTTACAGATTTGCCATCTGCTGTGACT-- 4933

Qy     1136 sProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSe 1156
                                     :::      :::|||   ::::::||||||:::
Db     4934 -------------------------GCTGACCAAGAATCCCACAAGGTAGAGACTTTA-- 4966

Qy     1156 rAsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSe 1176
                                     |||      |||      :::|||      :::
Db     4967 ----------------------CCGTATGTGCCTGAACCGGTTAAAGTGGCAATTGCAGA 5004

Qy     1176 rSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLy 1196
             :::   :::|||      ::::::                  ::::::|||:::
Db     5005 AAATCTGTTGGATGTAATTAAAGACACCAGAAGTAAGGAAGCAACTCCCGTGGCAGCAGG 5064

Qy     1196 sLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGl 1216
             ::::::            |||   :::   :::|||                     |||
Db     5065 TGAGGCTGGTGATGAGGACGGAGCAGTGATAGTCTCAAAGGCTGCACATTCGTCCAGGCT 5124

Qy     1216 yArgLysLysThrProValThrGluGlnGluGluLysLeu------------------Gl 1230
                     ||||||   |||   :::|||   :::
Db     5125 GACAAACTCTACACCGAAGACTGTTAAGGAACCACGTGCAGAGACTGTAAATACCAGCCA 5184

Qy     1230 yMetAspAspLeuThr------------------------------------------Ly 1236
              ||||||:::
Db     5185 GAGTGATGACATGGTTTCTTCTAGAACTCTCACAAGAAGGCAGCATGCCCTAAGCCTGAA 5244

Qy     1236 sLeuValGlnGluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAl 1256
             :::         ||||||:::|||               :::||||||
Db     5245 TGTCACATCAGAACAAGAGCCTTCAGCAGTTGCCACTCCTAAGAAGAGAACTAGAAAAAT 5304

Qy     1256 aSerGlu---------------------------SerAspGluGlnGlnTrpProGlu-- 1266
                |||                           ||||||   :::      ||||||
Db     5305 TAAAGAAACTCCTGAGTCTTCTGAAAGGACCTGTTCTGACCTAAAAGTAGCACCTGAGAA 5364

Qy     1267 -------------------------------GluLysArgLeuLysGluAspIle----- 1274
                                           :::|||   |||:::|||:::
Db     5365 CCAACTGACAGCTCAGAATCCTCCCGCTCCTAGGAGAAGAAAGAAGAAGGACGTTAGCCA 5424

Qy     1275 -------------------------LeuGluAsnGluAspGlu---GlnAsnSerProPr 1285
                                     :::|||   |||   |||   |||   :::|||
Db     5425 AGGCACACTGCCAAGTTCTGGTGCTGTGGAGCCGGAGCCGGAACCTCAGGGTACGCCGGG 5484

Qy     1285 oLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGlyThrPro---LysGl 1304
             :::         |||   :::||||||:::|||            ||||||   :::
Db     5485 AAGACTGAGGCTGAGAACGCAGCCACCCGAGCCAGCAGCTGAAGAAACTCCTTCTAGAAC 5544

Qy     1304 uGluProThrMetLysThrSerLysLysGlySerLysLysLys----------------- 1318
             :::      :::   :::   :::||||||:::   ::::::
Db     5545 AAAAGTCAGGCTTTCATCTGTTAGAAAGGGAACCCCTAGAAGACTTAAGAAGTCTGTAGA 5604

Qy     1319 ----------------------------------SerGlyProProAla----------- 1323
                                                 |||         |||
Db     5605 AAATGGGCAAAGTATAGAAATTCTAGATGATCTCAAAGGGAGTGAGGCAGCAAGTCATGA 5664

Qy     1324 ----------------------------------ProGluGluGluGluGluGluGluAr 1332
                                                 :::   :::         |||   ::
Db     5665 CGGGACTGTCACAGAGCTGAGGAATGCCAATTTAGAAGATACTCAGAATATGGAGTATAA 5724

Qy     1332 gGlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgValSerArgArgAl 1352
            :|||         :::::::::|||:::               |||:::   ||||||         |||
Db     5725 ACAAGATGAACACAGTGACCAGCAACCGCCTCTAAAACGAAAGAGGGTCAGAGAGAGAGA 5784

Qy     1352 aGlnGlnArgAlaGluSerProGluSerSerAlaIleGluSerThrGln----------- 1368
                     :::               |||         ::::::|||:::|||
Db     5785 AGTTAGTGTGTCAAGTGTGACAGAAGAGCCAAAGCTTGACTCATCCCAGTTGCCTCTTCA 5844
```

```
Qy       1369 ------------------SerThrProGlnLysGlyArgGlyArgProSerLys 1380
                                 :::||||||:::|||   |||||||||||    |||
Db       5845 GACAGGACTCGATGTACCTGCCACCCCTAGGAAA---CGTGGTAGACCCAGGAAG 5896

RESULT 14
US-09-882-227-623
; Sequence 623, Application US/09882227
; Publication No. US20030158396A1
; GENERAL INFORMATION:
;  APPLICANT: Kleanthous, Harold
;  APPLICANT:  Al-Garawi, Amal
;  APPLICANT:  Miller, Charles
;  APPLICANT:  Tomb, Jean-Francois
;  APPLICANT:  Ooomen, Raymond P.
;  TITLE OF INVENTION: Identification of Polynucleotides
;  TITLE OF INVENTION:  Encoding No. US20030158396A1el Helicobacter Polypeptides in the
;  TITLE OF INVENTION:  Genome
;  FILE REFERENCE: 06132/047002
;  CURRENT APPLICATION NUMBER: US/09/882,227
;  CURRENT FILING DATE:  2001-06-15
;  PRIOR APPLICATION NUMBER: US 08/902,615
;  PRIOR FILING DATE: 1997-07-29
;  NUMBER OF SEQ ID NOS: 638
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 623
;   LENGTH: 5334
;   TYPE: DNA
;   ORGANISM: Helicobacter pylori
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (22)...(5250)
US-09-882-227-623

Alignment Scores:
Pred. No.:              1.25e-09       Length:        5334
Score:                  225.00         Matches:       285
Percent Similarity:     32.49%         Conservative:  217
Best Local Similarity:  18.45%         Mismatches:    523
Query Match:            3.13%          Indels:        520
DB:                     12             Gaps:          71

US-09-512-581B-2 (1-1391) x US-09-882-227-623 (1-5334)

Qy          4 SerLysThrArgThrAsn---AspGlyLysIleThrTyrProProGlyValLysGluIle 22
              :::::|||:::||||||  ||||||        :::                :::||||||
Db        955 ACAGAAACTCAAACCAATTTGATGGAGACAAGTCA---------------GAAGAAATA 999

Qy         23 SerAspLysIleSerLysGluGluMetValArgArgLeuLys------------------ 36
              :::|||        :::    :::|||:::::::::          |||
Db       1000 ACTGACGACTCTAACGATCAAGAGATTATCAAAGGAAGCAAAAAGAAATATATTATTGGT 1059

Qy         37 ---------------------------------------------MetValValLysThr 41
                                                           :::
Db       1060 GGCATTGTAGTCGCTGTTCTTATCGTGATTATTTATTTCTAGAAGCATTTTCACTAC 1119

Qy         42 PheMetAspMetAspGlnAspSer---------GluGluGluLysGluLeuTyrLeuAsn 58
              ||||||   ::::::       |||              :::::::::    ||||||:::|||
Db       1120 TTCATGCCTTTGGAAGATAAAAGCTCTCGTTTTAGCAAAGACAGGAATCTTTATGTCAAT 1179

Qy         59 LeuAlaLeuHisLeuAlaSerAspPhe------PheLeuLysHisProGlyLysAspVal 76
                   :::    :::      ::::::          ||||||          :::
Db       1180 GATGAAATCCAAATAAGGCAAGAGTATAACCGATTGCTGAAAGAACGGAATGAAAAAGGC 1239

Qy         77 ArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaPro 96
                 ::::::
Db       1240 AATATGATCGAT------------------------------------------------ 1251

Qy         97 TyrThrSerProAspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGly 116
                                    |||::::::|||
Db       1252 --------------------AAGAATCTTTTCTTC------------------------- 1266

Qy        117 LeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAla 136
                                 |||             |||   ||||||        |||||||||
Db       1267 ---------AATGACGATCCCAATAGAACCTTATACAACTATTTG------AATATTGCA 1311

Qy        137 TrpValLysSerTyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGln 156
                                       |||:::||||||    |||    :::
Db       1312 --------------------------GAAATTGAGGACAAAAACCCGTTG--------- 1335

Qy        157 LeuTyrArgThrLeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHis 176
                    |||      :::      |||:::||||||   |||    :::      :::
Db       1336 ------AGAGCCTTTTATGAATGTATTAGTAATGGTGGCAACTATGAAGAATGTTTGAAG 1389

Qy        177 MetValAspLeuMetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeu 196
              ::::::
Db       1390 CTTATC------------------------------------------------------ 1395

Qy        197 AspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAsp 216
                                                |||   |||      |||
Db       1396 --------------------------AAAGACAAAAAACTTCAAGATCAG--------- 1419

Qy        217 LeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhe 236
                             :::|||:::|||   :::|||
Db       1420 ---------------ATGAAAAAGACTCTAGAGGCT----------------------- 1440

Qy        237 PheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPhe 256
              :::|||      :::         ||||||
Db       1441 TATAACGACTGCATCAAAAATGCCAAAACT----------------------------- 1470

Qy        257 AspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGln 276

Db       1470 ------------------------------------------------------------ 1470

Qy        277 LeuGluPheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeu 296
                                        |||||||||::::::   :::   |||:::
Db       1471 --------------------------GAAGAAGAAAGGATCAAGTGTTTAGATTTAATC 1503

Qy        297 AlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGln 316
                               ||||||     |||                |||   |||
Db       1504 -----------------AAAGATGAAAACCTA---------AAAAAAAGCTTA------ 1530

Qy        317 CysTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPhe 336
                               |||      |||   :::   |||:::|||:::|||
Db       1531 ---------------CTGAACCAACAAAAAGTTCAAGTGGCGCTAGATTGTTTGAAAAAC 1575

Qy        337 Ala------------------SerHisCys------LeuMetAsnHisPro--------- 345
              |||                  :::   |||      |||:::|||   |||
Db       1576 GCTAAAACCGATGAAGAACGAAACGAGTGCCTAAAACTCATAAATGACCCTGAGATTAGA 1635

Qy        346 --------------------AspLeuAlaLysAspLeuThrGluTyrLeuLysValArg 358
                                  :::|||   |||:::|||   ||||||
Db       1636 GAGAAATTCCGTAAGGAATTAGAGCTTCAAAAAGAGCTTCAAGAGTAT----------- 1683

Qy        359 SerHisAspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAla 378
                                         |||   |||   :::      |||   |||
Db       1684 --------------------------AAGGATTGTATCAAAAACGCCAAAACAGAAGCT 1716

Qy        379 LysLysAspIleLeuLeu------ValAsnAspHisLeuLeuAsnPheValArgGluArg 396
              :::|||:::      |||      ::::::          :::      :::::::::::
Db       1717 GAGAAAAACAAATGCTTGAAAGGCTTGTCTAAAGAAGCTATAGAGAGATTGAAACAGCAA 1776

Qy        397 ThrLeuAsp---------------LysArgTrpArgValArgLysGluAlaMetMetGly 411
                 ||||||               |||          |||   |||   :::
Db       1777 GCGCTAGATTGTTTGAAAAACGCTAAAACCGATGAAGAACGAAACGAGTGCTTGAAAAAT 1836

Qy        412 LeuAlaGlnIleTyrLysLysTyrAlaLeuGln--------------------------- 422
              :::   |||      :::|||   |||
Db       1837 ATTCCCCAAGACTTGCAAAAAGAACTATTAGCTGATATGAGCGTCAAGGCTTACAAGGAT 1896

Qy        423 ------SerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu 440
                    |||   |||      :::      ||||||               :::||||||
Db       1897 TGCGTATCAAAAGCTAGAAATGAAAAAGAGAAACAAGAATGC---------GAGAAATTG 1947

Qy        441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
              |||               :::      ::::::   ::::::|||
Db       1948 CTCACGCCTGAAGCGAGGAAAAAGTTAGAACAACAGGTTCTA------------------ 1989
```

```
Qy     461 GlnTyrMetValProHisAsnLeuGluThr---ThrGluArgMetLysCysLeuTyrTyr 479
                             |||   :::|||      ||||||   |||||||||
Db    1990 ------GATTGTTTGAAAAACGCTAAAACCGATGAAGAACGAAAAAAGTGTTTGAAAGAT 2043

Qy     480 LeuTyrAlaThrLeu------AspLeuAsnAlaValLysAlaLeuAsn------------ 493
           |||         |||      |||:::   |||   ::::::|||
Db    2044 CTCCCTAAAGACTTACAAAGCGATATTCTAGCCAAAGAGAGCCTGAAAGCTTATAAAGAC 2103

Qy     494 ---------------------------GluMetTrpLysCysGlnAsnLeuLeu------ 502
                                      |||      :::|||:::   ||||||
Db    2104 TGCGTATCTCAAGCCAAAACCGAAGCTGAGAAAAAAGAATGCGAGAAATTACTCACCCCT 2163

Qy     503 ------------------------------ArgHisGlnValLysAspLeuLeuAspLeu 512
                                         :::      ||||||      ||||||
Db    2164 GAAGCGAAAAAACTTTTAGAAGAAGAAGCCAAAGAGAGCGTTAAGGCTTATTTGGATTGC 2223

Qy     513 IleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysValMetValIle 532
           :::   |||   ||||||:::|||      |||            |||::::::
Db    2224 GTATCTCAAGCCAAAACCGAAGCTGAGAAAAAAGAATGC---GAGAAATTGCTCACC--- 2277

Qy     533 ThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnVal 552
                       |||:::      |||      :::      ||||||      :::
Db    2278 ------------CCTGAAGCGAAAAAAAAGTTAGAAGAAGCTAAAAAAAGCGTTAAAGCT 2325

Qy     553 LeuGluAsp------------------AspGluLysIleArgLysGlnLeuGluValLeu 566
                 |||                  :::||||||   :::|||:::   |||   |||
Db    2326 TACTTGGATTGCGTATCAAGAGCTAGGAATGAAAAAGAGAAAAAAGAATGCGAGAAATTG 2385

Qy     567 ValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLys 586
           ::::::|||                                       :::   |||
Db    2386 CTCACCCCTGAAGCGAAA---AAACTTTTAGAGCAACAAGCACTAGATTGTTGAAAAAC 2442

Qy     587 LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGlu 606
                       |||:::               ::::::   |||      |||   :::
Db    2443 GCTAAAACCGATAAAGAACGAAAAAAGTGTTTGAAAGATCTCCCTAAAGACTTGCAGAAA 2502

Qy     607 ArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsn 626
           ::::::            :::      ||||||:::   |||   :::      |||:::
Db    2503 AAGGTT------------TTAGCTAAAGAAAGCGTTAAAGCTTACTTGGATTGCGTATCT 2550

Qy     627 LysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIle 646
           ::::::                     |||   |||   :::   |||   :::|||
Db    2551 CAAGCCAAAACTGAAGCTGAGAAAAAAGAATGCGAGAAATTACTCACCCCTGAAGCG--- 2607

Qy     647 ArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSer 666
           |||      ||||||      |||
Db    2608 AGAAAACTTTTAGAAGAAGCTAAA------------------------------------ 2631

Qy     667 AlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGlu 686
              ::::::   ::::::   |||   |||                     |||::::::
Db    2632 ---AAAAGCGTTAAGGCTTAYTTGGATTGC---------------------GTATCTCAA 2667

Qy     687 AlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHisIle 706
           |||      :::      |||            |||:::            |||
Db    2668 GCCAAAACTGAAGCTGAGAAAAAAGAATGCGAGAAATTACTCACC------CCTGAAGCG 2721

Qy     707 ArgSerAlaLeuLeuProValLeuHisHisLysSerLysLysGlyProProArgGlnAla 726
           |||      |||                                          :::|||
Db    2722 AGAAAACTCTTA---------------------------------------GAAGAAGCT 2742

Qy     727 LysTyrAlaIle------------HisCysIleHisAlaIlePheSerSerLysGluThr 742
           |||   ::::::                                 :::   ||||||
Db    2743 AAAGAGAGCGTTAAAGCTTATAAAGACTGCGTATCAAAAGCTAGGAATGAAAAAGAGAAA 2802

Qy     743 GlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeuGluHis 762
           :::                                                   |||
Db    2803 AAA---------------------------------------------GAATGCGAGAAA 2817

Qy     763 LeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAla 782
           |||:::||||||                           |||         |||
Db    2818 TTACTCACGCCT---------------GAAGCGAAAAAACTTTAGAGCAACAA------ 2856

Qy     783 AlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsnAspArg 802
                                            |||   |||   |||   |||   :::
Db    2857 ---------------------------------GTGCTAGATTGTTTGAAAAACGCTAAA 2883

Qy     803 LeuProGlyLysLysThrThrLysLeuTrpValProAsp------GluGluValSerPro 820
                       |||      |||      |||   |||      :::::::::
Db    2884 ACCGAAGCTGATAAA------AAAAGGTGTGTCAAAGATCTCCCTAAAGACTTGCAGAAA 2937

Qy     821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLys 840
           :::   :::   |||   :::::::::|||      :::      :::         :::
Db    2938 AAGGTTTTAGCTAAA---GAGAGCGTTAAGGCTTATTTGGACTGCGTATCAAGAGCTAGG 2994

Qy     841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
           |||         |||
Db    2995 AATGAAAAAGAGAAAAAA------------------------------------------ 3012

Qy     861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
                          |||      |||:::      |||:::      :::|||
Db    3013 ---------------GAATGCGAGAAATTGCTCACCCCTGAAGCGAAAAAACTTTTAGAA 3057

Qy     881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
           |||   ::::::               |||   :::   |||      :::
Db    3058 GAAGCCAAAGAGAGTCTT------AAAGCTTATAAAGACTGCCTCTCTCAAGCTAGAAAT 3111

Qy     901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
              |||:::   :::   |||         :::      |||            |||:::
Db    3112 GAAGAAGAAAGGAGAGCTTGCGAGAAACTACTCACGCCTGAA---------GCGAGAAAA 3162

Qy     921 ValPheAlaGlnLysLeuHisLysGlyLeu---------------------SerArgLeu 933
           :::      |||::::::   |||   :::                     ||||||
Db    3163 CTCTTAGAGCAAGAAGTTAAGAAAAGCATTAAGGCTTATTTGGACTGCGTATCAAGAGCT 3222

Qy     934 ArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGlu 953
           |||         |||         |||                  :::      |||
Db    3223 AGGAATGAAAAAGAGAAAAAAGAATGCGAGAAATTACTCACGCCTGAAGCGAGAAAATTT 3282

Qy     954 ArgArgAlaHisAlaArgGlnCysLeuValLys------AsnIleAsnValArgArgGlu 971
                                ||||||   |||                  |||:::
Db    3283 TTAGCGAAGCAAGTGCTAAATTGTTTGGAAAAAGCTGGAAATGAAGAAGAAAGAAAAGCA 3342

Qy     972 TyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrVal 991
              ||||||                                    ||||||:::   :::
Db    3343 TGTCTTAAAAAT---------------------------------CTCCCTAAAGACTTA 3369

Qy     992 ValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIle 1011
                          ::::::||||||
Db    3370 CAGGAA---------AATATTTTAGCTAAA------------------------------ 3390

Qy    1012 GluGlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLys 1031
           |||   ||||||      |||:::||||||                  |||:::
Db    3391 GAGAGTCTTAAAGCTTATAAAGACTGCCTC------------------TCTCAAGCTAGA 3432

Qy    1032 AsnGluAsnAsnSerHisAlaPhe---------------------IleArgLysMetVal 1044
           ||||||            |||                           ||||||::::::
Db    3433 AATGAAGAAGAAAGGAGAGCTTGCGAGAAACTACTCACGCCTGAAGCGAGAAAACTCTTA 3492

Qy    1045 GluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLys 1064
           |||                                    :::   |||   :::   |||
Db    3493 GAG---------------------------------CAAGAAGTTAAGAAAAGCGTTAAG 3519

Qy    1065 LeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSer 1084
              |||      |||                           |||
Db    3520 GCTTATTTGGACTGCGTATCAAGAGCTAGGAATGAAAAAGAGAAAAAAGAATGCGAGAAA 3579

Qy    1085 LeuGluSerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsn 1104
           |||   :::|||:::            |||:::            |||   ||||||
Db    3580 TTACTCACGCCTGAAGCGAGAAAATTTTTAGCGAAAGAACTCCAACAAAAAGATAAAGCG 3639

Qy    1105 PheSerAsn---ThrLysAsnTyrLeuPro------------------------------ 1113
                 :::      ||||||   |||
Db    3640 ATCAAAGATTGCTTGAAAAACGCCGATCCTAACGACAGAGCGGCTATCATGAAGTGTTTG 3699

Qy    1114 ---------------ProGluMetLysSerPhePheThrProGlyLysProLysThrThr 1128
```

```
                                |||       :::              :::   |||
Db       3700 GATGGTTTGAGCGATGAAGAGAAGCTCAAATACCTGCAAGAAGCTAGAGAAAAGGCTGTT 3759

Qy       1129 AsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLys 1148
                          |||:::   |||                    |||
Db       3760 GCGGATTGTTTGGCTATGGCTAAAACCGATGAAGAAAAAAGGAAATGCCAAAACCTTTAT 3819

Qy       1149 SerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGly 1168
               |||     :::::  :::     |||      :::   :::      :::
Db       3820 AGCGATTTGATCCAAGAAATCCAAAATAAAAGGACACAAAACAAACAAAATCAATTGAGT 3879

Qy       1169 ArgIleLysGlyArgLeuAspSerSerGlu---MetAspHisSerGluAsn------Glu 1185
              :::   :::            :::||||||  :::|||:::     ::::::      :::
Db       3880 AAAACAGAAAGGTTGCATCAAGCAAGCGAGTGCTTGGATAACTTAGATGACCCTACTGAT 3939

Qy       1186 AspTyrThrMetSerSerProLeuProGlyLysLysSerAspLysArg----------- 1201
                       :::           |||   |||             :::|||
Db       3940 CAAGAGGCCATAGAGCAATGTTTAGAGGGCTTGAGCGATAGTGAAAGGGCGCTAATTCTA 3999

Qy       1202 -----------------AspAspSerAspLeuValArgSerGluLeuGluLysProArg 1215
                                |||:::   ||||||:::   |||:::|||          :::
Db       4000 GGAATTAAACGACAAGCTGATGAAGTGGATCTGATTTATAGCGATCTAAGAAACCGTAAA 4059

Qy       1216 --------------GlyArgLysLysThrProValThrGluGlnGluGluLysLeuGly 1230
                                   |||      |||:::         :::   |||   |||
Db       4060 ACCTTTGATAACATGGCGGCTAAAGGTTATCCATTGTTACCAATGGATTTCAAAAATGGC 4119

Qy       1231 MetAsp--------------------------AspLeuThrLysLeuValGlnGluGln 1241
                  |||                            |||   |||:::         :::
Db       4120 GGCGATATTGCCACTATTAACGCCACTAATGTTGATGCGGACAAAATAGCTAGCGATAAT 4179

Qy       1242 LysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp--- 1260
                                                  :::   :::   |||   |||
Db       4180 CCTATT-----------------------------TATGCTTCCATAGAGCCTGATATT 4209

Qy       1261 GluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGlu 1280
                 :::|||:::      ||||||   :::|||:::      ||||||
Db       4210 GCCAAGCAATACGAAACAGAAAAAACCATTAAGGATAAGAATTTAGAA------------ 4257

Qy       1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
                                               :::   |||      |||||||||
Db       4258 -----------------------------GCTAAATTAGCTAAGGCTTTAGGTGGCAAT 4287

Qy       1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLysSerGly 1320
                    |||::::::                        |||||||||   :::   :::
Db       4288 ------AAAAAAGATGACGATAAAGAAAAAAGTAAAAAATCCACAGCAGAAGCTAAAGCA 4341

Qy       1321 ProProAlaProGluGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys 1340
                             ::::::::::     |||                  :::|||
Db       4342 GAAAACAATAAGATAGACAAAGATGTCGCAGAAACTGCCAAGAATATCAGTGAAATCGCT 4401

Qy       1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGlu------Ser 1358
                 |||:::|||:::   :::                  :::
Db       4402 CTTAAGAACAAAAAAGAAAAGAGTGGGGAATTTGTAGATGAAAATGGTAATCCCATTGAT 4461

Qy       1359 ProGluSerSerAlaIleGluSerThrGlnSerThrProGlnLys-----------Gly 1374
                 :::      |||   :::      :::::::::|||      |||            |||
Db       4462 GACAAAAAGAAAGCAGAAAAACAAGATGAAACAAGCCCTGTCAAACAGGCCTTTATAGGC 4521

Qy       1375 ArgGlyArgProSer 1379
              :::      |||:::
Db       4522 AAGAGTGATCCCACA 4536
```

RESULT 15
US-10-171-581-49
; Sequence 49, Application US/10171581
; Publication No. US20030104426A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: Linsley, Peter
; APPLICANT: Mao Mao
; TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
; FILE REFERENCE: 9301-157-999
; CURRENT APPLICATION NUMBER: US/10/171,581
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,914
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 366
; SEQ ID NO 49
; LENGTH: 3727
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: D21262
; DATABASE ENTRY DATE: 2001-06-18
US-10-171-581-49

```
Alignment Scores:
Pred. No.:                 4.26e-08        Length:         3727
Score:                     205.50          Matches:        141
Percent Similarity:        33.64%          Conservative:   77
Best Local Similarity:     21.76%          Mismatches:     256
Query Match:               2.86%           Indels:         175
DB:                        14              Gaps:           26
```

US-09-512-581B-2 (1-1391) x US-10-171-581-49 (1-3727)

```
Qy        858 LeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArg 877
              |||||||||                         ::::::           |||      |||
Db        115 CTACACAGC------------------------AGGATGCCAATGCCTCTTCCCTCTTAG 150

Qy        878 LeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGlu 897
                    :::|||:::||||||::::::   :::   :::|||            |||    :::
Db        151 ACATCTATAGCTTCTGGCTCAAGTCTGCCAAGG---TCCCAGAGCGAAAGTTACAGGCAA 207

Qy        898 IleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGln 917
                       ::::::|||:::   |||      |||
Db        208 ---------ATGGACCAGTGG---CTAAGAAAGCTAAGAAGAAGGCCTCATCCAGTGACA 255

Qy        918 ValArgGlnValPhe---AlaGlnLysLeuHisLysGlyLeuSerArgLeuArg------ 934
              ||||||             ::::::          |||||||||   :::   |||
Db        256 GTGAGGACAGCAGCGAGGAGGAGGAGGAAGTTCAAGGGCCTCCAGCAAAGAAGGCTGCTG 315

Qy        935 ---LeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGlu 953
                 ||||||   |||       ::::::|||      |||          |||         :::
Db        316 TACCTGCCAAGCGAG------TCGGTCTGCCTCCTGGGAAGGCTGCAGCCAAAGCATCAG 369

Qy        954 ArgArgAlaHisAlaArgGlnCysLeuVal--------------------LysAsnIle 966
              |||   |||      :::       ::::::                        :::|||
Db        370 AGAGTAGCAGCAGTGAAGAGTCCAGTGATGATGATGATGAGGAGGACCAAAGAAACAGC 429

Qy        967 AsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeu 986
                    |||||||||          :::                   |||||||||
Db        430 CTGTCCAGAAGGGAGTTAAGCCCCAAGCCAAGG------CAGCCAAAGCTCCTCCTAAGA 483

Qy        987 LeuProGluTyrValVal-----------------ProTyrThrIleHisLeuLeuAla 1000
                 |||   ::::::                        |||         |||
Db        484 AGGCCAAGAGCTCTGATTCTGATTCTGACTCAAGCTCCGAGGATGAGCCACCAAAGAACC 543

Qy       1001 HisAspProAspTyr------------ValLysValGlnAspIleGluGlnLeuLysAsp 1016
                    |||                  ||||||   |||                |||
Db        544 AGAAG-CCAAAGATAACACCTGTGACAGTTAAAGCTCAGACTAAAGCCCCTCCCAAACCA 602

Qy       1017 ValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSer 1036
                 :::                                      |||      |||
Db        603 GCTCGAGCAGCA-------------------------CCTAAAATAGCCAATGGTAAA 635

Qy       1037 HisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPro 1056
                 |||                      :::         :::   |||::::::
Db        636 GCAGCCAGTAGCAGCAGTAGCAGCAGCAGCAGCAGTAGCAGTGATGACTCAGAGGAGGAG 695

Qy       1057 AspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIle 1076
                    |||          :::|||::::::|||:::            |||             :::
Db        696 AAGGCAGCAGCCACCCCCAAGAAGGTCTGGACCATAACTTCTGTCAGGGCAGAGACTGTA 755

Qy       1077 MetSerLysSerThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArg 1096
```

```
                      |||                             |||   ||||||          |||
Db          756 CCTAAAAGCAAGTTGTG--------------GCCAAGGCCCCAGTGAAAGCAGCTACC 800

Qy         1097 PhePheThr------------------------------------------------- 1099
                      |||
Db          801 ACCCCTACCCGGAAGAGTTCTAGCAGTGAGGATTCCTCCAGTGACGAGGAAGAGGAGCAA 860

Qy         1100 ---GlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSer 1118
                   :::|||   ||||||                     :::   :::||||||
Db          861 AAAAAACCCATGAAAAATAAACCAGGTCCCTACAGTTCAGTCCCCCCG----------- 908

Qy         1119 PhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeu 1138
                         |||      |||               ||||||            |||
Db          909 ---------CCTTCTGCTCCCCCACCAAAGAAGTCTCTGGGA---ACCCAGCCTCCCAAG 956

Qy         1139 SerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAla 1158
                   :::      ||||||         :::||||||      |||      |||:::
Db          957 AAGGCTGTGGAGAAGCAGCAGCCTGTGGAAAGCAGT------GAAGACAGCAGTGATGAG 1010

Qy         1159 SerSerSerSerAsn--------------------------------------------- 1163
                |||   ||||||:::
Db         1011 TCTGATTCAAGTTCTGAAGAAGAGAAGAAACCCCCAACTAAGGCAGTAGTCTCTAAAGCA 1070

Qy         1164 -----------ProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMet 1179
                            |||   :::            :::         |||||||||:::
Db         1071 ACCACTAAACCACCTCCAGCAAAGAAAGCAGCAGAGAGCTCTTCAGACAGCTCAGACTCT 1130

Qy         1180 AspHisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAsp 1199
                |||   ||||||   :::|||         |||   |||               |||:::
Db         1131 GACAGCTCTGAG---GATGATGAAGCTCCTTCTAAGCCAGCTGGTACCACCAAGAATTCT 1187

Qy         1200 LysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLys 1219
                            :::         :::|||         ||||||            |||
Db         1188 TCAAATAAGCCAGCTGTCACCACCAAGTCACCTGCAGTGAAGCCAGCTGCAGCCCCCAAG 1247

Qy         1220 ThrProVal---------------------------------------------ThrGlu 1224
                   ||||||                                             :::|||
Db         1248 CAACCTGTGGGCGGTGGCCAGAAGCTTCTGACGAGAAAGGCTGACAGCAGCTCCAGTGAG 1307

Qy         1225 GlnGluGluLysLeuGlyMetAspAspLeuThr---LysLeuValGlnGluGlnLysPro 1243
                :::|||               ::::::      |||   |||:::|||         ||||||
Db         1308 GAAGAGAGCAGCTCCAGTGAGGAGGAGAAGACAAAGAAGATGGTGGCCACCACTAAGCCC 1367

Qy         1244 LysGlySerGlnArgSer-----------------------ArgLysArgGlyHisThr 1255
                |||   :::   ::::::                                    :::|||
Db         1368 AAGGCGACTGCCAAAGCAGCTCTACCTCTGCCTGCCAAGCAGGCTCCTCAGGGTAGTAGG 1427

Qy         1256 AlaSerGluSerAspGluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeu 1275
                   |||   ||||||                                      |||
Db         1428 GACAGCAGCTCTGAT--------------------------------TCAGACAGCTCC 1454

Qy         1276 GluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgProProLys 1295
                   :::|||:::|||:::            |||         :::   :::|||   |||
Db         1455 AGCAGTGAGGAGGAGGAAGAGAAGACATCTAAGTCTGCAGTTAAGAAGAAGCCACAGAAG 1514

Qy         1296 ProLeuGlyGlyGlyThrPro--------------------------------------- 1302
                      ||||||      |||
Db         1515 GTAGCAGGAGGTGCAGCCCCTTCCAAGCCAGCCTCTGCAAAGAAAGGAAAGGCTGAGAGC 1574

Qy         1303 ---------------------------LysGluGluProThrMetLysThrSerLysLys 1313
                                              ||||||         |||         |||
Db         1575 AGCAACAGTTCTTCTTCTGATGACTCCAGTGAGGAAGAGGAAGAGAAGCTCAAGGGCAAG 1634

Qy         1314 GlySerLysLysLysSerGlyProProAla----------------------------- 1323
                ||||||   :::         |||   |||
Db         1635 GGCTCTCCAAGACCACAAGCCCCCAAGGCCAATGGCACCTCTGCACTGACTGCCCAGAAT 1694

Qy         1324 ------------------ProGluGluGluGluGluGluGluArgGlnSerGlyAsnThr 1337
                                     ||||||||||||||||||::::::::::::
Db         1695 GGAAAAGCAGCTAAGAACAGTGAGGAGGAGGAAGAAGAAAAGAAAAAGGCGGCAGTGGTA 1754

Qy         1338 GluGlnLysSerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGlu 1357
                      ||||||   |||   :::   |||            |||   :::   ||||||
Db         1755 GTTTCCAAATCAGGTTCATTAAAGAAGCGGAAGCAGAATGAGGCTGCCAAGGAGGCAGAG 1814

Qy         1358 SerProGluSerSerAlaIleGlu---SerThrGlnSerThrProGlnLysGlyArgGly 1376
                :::|||::::::      |||:::      |||   :::|||      |||   :::
Db         1815 ACTCCTCAGGCCAAGAAGATAAAGCTTCAGACCCCTAACACATTTCCAAAAAGGAAGAAA 1874

Qy         1377 ArgProSerLysThrProSerPro 1384
                         :::      ||||||
Db         1875 GGAGAAAAAAGGGCATCATCCCCA 1898
```

Search completed: September 25, 2003, 04:38:19
Job time : 1030 secs

NPA


```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         September 24, 2003, 20:26:40 ; Search time 167 Seconds
                                          (without alignments)
                                          3676.431 Million cell updates/sec

Title:          US-09-512-581B-2
Perfect score:  7193
Sequence:       1 MAHSKTRTNDGKITYPPGVK..........QKGRGRPSKTPSPSQPKKNV 1391

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters:      1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09512581/runat_23092003_163547_5508/app_query.fasta_1.1543
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09512581_@CGN_1_1_85_@runat_23092003_163547_5508 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                2:  /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                6:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	349	4.9	530	3	US-09-328-111-682	Sequence 682, App
2	231	3.2	8503	4	US-09-620-312D-130	Sequence 130, App
3	229	3.2	8257	4	US-09-595-684B-30	Sequence 30, Appl
4	207	2.9	6755	3	US-08-931-999-4	Sequence 4, Appli
5	201.5	2.8	10136	1	US-08-353-700-2	Sequence 2, Appli
6	201.5	2.8	10136	5	PCT-US95-16216-2	Sequence 2, Appli
7	197	2.7	8789	1	US-08-328-254-5	Sequence 5, Appli
8	191	2.7	10320	4	US-09-091-501B-9	Sequence 9, Appli
9	190	2.6	30549	4	US-09-134-001C-322	Sequence 322, App
10	189.5	2.6	4868	1	US-08-139-937-12	Sequence 12, Appl
11	189.5	2.6	4868	5	PCT-US93-11310-12	Sequence 12, Appl
c 12	184.5	2.6	19307	3	US-08-836-022A-10	Sequence 10, Appl
c 13	184.5	2.6	19307	3	US-09-427-048A-10	Sequence 10, Appl
14	182.5	2.5	6306	1	US-08-195-487-3	Sequence 3, Appli
15	182.5	2.5	6306	5	PCT-US93-06160-3	Sequence 3, Appli
16	181	2.5	6605	1	US-08-769-309A-4	Sequence 4, Appli
17	181	2.5	6605	3	US-08-994-570-4	Sequence 4, Appli
18	181	2.5	6608	4	US-09-220-132-58	Sequence 58, Appl
19	180.5	2.5	13977	4	US-09-484-970B-60	Sequence 60, Appl
20	180	2.5	8351	1	US-08-198-446B-14	Sequence 14, Appl
21	180	2.5	8351	2	US-08-870-693-14	Sequence 14, Appl
22	179.5	2.5	6306	1	US-08-466-390-3	Sequence 3, Appli
23	179.5	2.5	6306	1	US-08-470-950-3	Sequence 3, Appli
24	179.5	2.5	6306	1	US-08-467-781-3	Sequence 3, Appli
25	179.5	2.5	6306	2	US-08-483-924-3	Sequence 3, Appli
26	179.5	2.5	7453	4	US-09-620-312D-248	Sequence 248, App
27	178	2.5	6152	3	US-08-973-462-1	Sequence 1, Appli
28	177	2.5	6921	4	US-09-643-597-117	Sequence 117, App
29	177	2.5	6921	4	US-09-480-884A-117	Sequence 117, App
30	177	2.5	6921	4	US-09-542-615A-117	Sequence 117, App
31	177	2.5	6921	4	US-09-606-421B-117	Sequence 117, App
32	176.5	2.5	7501	4	US-09-620-312D-249	Sequence 249, App
33	176	2.4	4716	2	US-08-290-731C-1	Sequence 1, Appli
34	175.5	2.4	6775	4	US-09-620-312D-289	Sequence 289, App
35	174.5	2.4	6519	1	US-08-588-985-1	Sequence 1, Appli
36	174.5	2.4	6519	1	US-08-971-988-1	Sequence 1, Appli
37	174	2.4	5361	3	US-08-973-462-2	Sequence 2, Appli
38	173.5	2.4	6773	4	US-09-166-350-27	Sequence 27, Appl
39	171	2.4	580073	4	US-08-545-528D-1	Sequence 1, Appli
40	170	2.4	9370	1	US-08-320-559-27	Sequence 27, Appl
41	170	2.4	9370	3	US-08-545-860D-27	Sequence 27, Appl
42	170	2.4	9370	5	PCT-US94-04496-27	Sequence 27, Appl
43	170	2.4	9391	1	US-08-320-559-25	Sequence 25, Appl
44	170	2.4	9391	3	US-08-545-860D-25	Sequence 25, Appl
45	170	2.4	9391	5	PCT-US94-04496-25	Sequence 25, Appl

ALIGNMENTS

```
RESULT 1
US-09-328-111-682/c
; Sequence 682, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
;  APPLICANT: Endege, Wilson O.
;  APPLICANT: Steinmann, Kathleen E.
;  APPLICANT: Astle, Jon H.
;  APPLICANT: Burgess, Christopher C.
;  APPLICANT: Bushnell, Steven E.
;  APPLICANT: Carroll III, Eddie
;  APPLICANT: Catino, Theodore J.
;  APPLICANT: Derti, Adnan
;  APPLICANT: Ford, Donna M.
;  APPLICANT: Lewis, Marcia E.
;  APPLICANT: Monahan, John E.
;  APPLICANT: Schlegel, Robert
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS
;  FILE REFERENCE: CCD-257 (US)
;  CURRENT APPLICATION NUMBER: US/09/328,111
;  CURRENT FILING DATE: 1999-06-08
;  EARLIER APPLICATION NUMBER: US 60/088,801
;  EARLIER FILING DATE: 1998-06-10
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 682
;   LENGTH: 530
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-328-111-682

Alignment Scores:
Pred. No.:             1.94e-24      Length:        530
Score:                 349.00        Matches:       67
Percent Similarity:    98.53%        Conservative:  0
```

```
Best Local Similarity:  98.53%        Mismatches:    1
Query Match:            4.85%         Indels:        0
DB:                     3             Gaps:          0

US-09-512-581B-2 (1-1391) x US-09-328-111-682 (1-530)

Qy          37 MetValValLysThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyr 56
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         213 ATGGTTGTGAAAACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTAT 154

Qy          57 LeuAsnLeuAlaLeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspVal 76
               ||||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db         153 TTAAACCTAGCTTTACATCTTGCTTCAGATTTTTTCCTCAAGCATCCTGATAAGGATGTT 94

Qy          77 ArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaPro 96
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          93 CGCTTACTGGTAGCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCT 34

Qy          97 TyrThrSerProAspLysLeuLys 104
               ||||||||||||||||||||||||
Db          33 TACACATCCCCTGATAAACTAAAG 10
```

RESULT 2
US-09-620-312D-130
; Sequence 130, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 130
; LENGTH: 8503
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (91)..(8082)
US-09-620-312D-130

```
Alignment Scores:
Pred. No.:              6.67e-11      Length:        8503
Score:                  231.00        Matches:       285
Percent Similarity:     35.41%        Conservative:  267
Best Local Similarity:  18.28%        Mismatches:    541
Query Match:            3.21%         Indels:        467
DB:                     4             Gaps:          67

US-09-512-581B-2 (1-1391) x US-09-620-312D-130 (1-8503)

Qy          17 ProGlyValLysGluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLys 36
               |||   |||   |||:::|||           :::   |||   ::::::::|||   :::
Db        1087 CCTTATGTTAATGAGGTATCA---------ACTGATGAAGCTCTCCTGAAAAGGTATAGA 1137

Qy          37 MetValValLysThrPheMetAspMetAspGlnAspSerGluGlu--------------- 51
                         |||      ||||||:::   :::      ||||||
Db        1138 ---------AAAGAAATAATGGATCTTAAAAAACAATTAGAGGAGGTTTCTTTAGAGACG 1188

Qy          52 ---------------GluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
                              ||||||:::               |||                |||
Db        1189 CGGGCTCAGGCAATGGAAAAAGACCAATTGGCCCAACTTTTGGAAGAA------AAAGAT 1242

Qy          67 PhePheLeuLysHisProGlyLysAspVal------------ArgLeuLeuValAlaCys 82
                         |||         :::   :::               |||:::||||||
Db        1243 TTGCTTCAGAAAGTACAGAATGAGAAAATTGAAAACTTAACACGGATGCTGGTG------ 1296

Qy          83 CysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLys 102
                                                          ||||||
Db        1297 ---------------------------------------------ACCTCTTCTTCC--- 1308

Qy         103 LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLysSer 122
                                  :::|||   |||               ::::::   |||:::
Db        1309 --------------------CTCACGTTGCAA------------CAGGAATTAAAGGCT 1335

Qy         123 ProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsn 142
                  :::      |||      :::   |||         |||      :::|||:::   |||
Db        1336 AAAAGAAAACGAAGAGTTACTTGGTGCCTTGGCAAAATTAACAAAATGAAGAACTCAAAC 1395

Qy         143 IleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPhe 162
                                                           |||         |||
Db        1396 ---------------------------------------------TATGCAGATCAATTT 1410

Qy         163 SerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSer 182
               ::::::      |||               |||   |||         ::::::|||:::
Db        1411 AATATACCAACAAAT---ATAACAACAAAAACACATAAGCTTTCTATAAATTTATTACGA 1467

Qy         183 SerIle---------IleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrVal 199
                  |||            :::|||            |||            |||||||||:::
Db        1468 GAAATTGATGAATCTGTCTGTTCAGAGTCTGATGTTTCAGTAACACTCTTGATACATTA 1527

Qy         200 Leu---ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAla 218
                     :::         ||||||   |||   ||||||::::::         :::
Db        1528 AGTGAGATAGAATGGAATCCAGCAACAAAGCTACTAAATCAGGAGAATATAGAA------ 1581

Qy         219 LysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsn 238

Db        1581 ------------------------------------------------------------ 1581

Qy         239 GlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeu 258
                                  :::   ::::::   |||                  :::|||
Db        1582 --------------------AGTGAGTTGAACTCACTTCGTGCTGACTATGATAATCTG 1620

Qy         259 IleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGlu 278
               :::|||               |||         |||   :::         ::::::|||
Db        1621 GTATTA---------------GACTATGAACAACTACGAACAGAAAAAAAAAAAATGGAA 1665

Qy         279 PheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLys 298
                  |||||||||      ::::::   :::      :::   ::::::         |||
Db        1666 TTGAAATTAAAAGAAAAGAATGATTTGGATGAATTTGAGGCTCTAGAAAGAAAAACTAAA 1725

Qy         299 MetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyr 318
                           ||||||   |||:::
Db        1726 ------------AAAGATCAAGAGATGCAA---------------------------- 1743

Qy         319 LeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSer 338

Db        1743 ------------------------------------------------------------ 1743

Qy         339 HisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArg 358
                     |||::::::               |||:::|||   ::::::   :::|||
Db        1744 ------CTAATTCATGAAATTTCGAACTTAAAGAATTTAGTTAAGCATCGAGAAGTATAT 1797
```

```
Qy       359 SerHisAspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAla 378
             :::    |||  |||      :::                              ::::::
Db      1798 AATCAAGATCTTGAGAATGAACTC------------------------------AGTTCA 1827

Qy       379 LysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeu 398
             |||   :::::|||          :::      :::     ::::::|||
Db      1828 AAAGTAGAGCTGCTTAGAGAAAAGGAAGACCAGATTAAGAAGCTACAGGAATACATAGAC 1887

Qy       399 AspLysArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLys 418
                ::::::            :::         |||   |||:::
Db      1888 TCTCAAAAG---------CTAGAAAATATAAAAATGGACTTGTCA--------------- 1923

Qy       419 TyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAsp 438
             |||:::|||:::|||                     ||||||      :::      :::
Db      1924 TACTCATTGGAAAGCATTGAAGACCCAAAACAAATGAAGCAGACTCTGTTTGATGCTGAA 1983

Qy       439 LysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIle 458
                :::                                       |||      :::
Db      1984 ACTGTA------------------------------------GCCCTTGATGCCAAGAGA 2007

Qy       459 PheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyr 478
             :::   ::::::           |||||||||     |||:::||||||     |||
Db      2008 GAATCAGCCTTTCTTAGAAGTGAAAATCTGGAGTTGAAGGAGAAAATGAAAGAACTTGCA 2067

Qy       479 TyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCys 498
                    |||     ::::::                     |||:::::
Db      2068 ACTACATACAAGCAAATGGAA---------------------AATGATATTCAGTTATAT 2106

Qy       499 GlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThr 518
             |||:::   |||      :::   |||      :::||||||   |||:::
Db      2107 CAAAGCCAATTGGAGGCAAAAAAGAAAATGCAAGTTGATCTGGAGAAAGAA--------- 2157

Qy       519 AspAlaSerValLysAlaIlePheSerLysValMetValIleThrArg------------ 534
                      :::::::::   |||:::::::::      :::|||
Db      2158 ---------TTACAATCTGCTTTTAATGAGATAACAAAACTCACCTCCCTTATAGATGGC 2208

Qy       535 ------------------------AsnLeuProAspProGlyLysAlaGlnAspPheMet 546
                                     ||||||   ||||||            |||
Db      2209 AAAGTTCCAAAAGATTTGCTCTGTAATTTGGAATTGGAAGGAAAGATTACTGATCTTCAG 2268

Qy       547 LysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeu 566
             |||:::      :::   :::|||::::::|||   :::|||:::::::::   :::|||
Db      2269 AAAGAACTAAATAAAGAAGTTGAAGAAAATGAAGCTTTGCGGGAAGAAGTC---ATTTTG 2325

Qy       567 ValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLys 586
             :::|||               |||               |||      :::   |||:::
Db      2326 CTTTCA---------GAATTGAAATCTTTACCTTCTGAAGTAGAAAGGCTGAGGAAAGAG 2376

Qy       587 LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGlu 606
             :::            |||      :::      |||   :::|||            :::
Db      2377 ATA------------CAAGACAAATCTGAAGAGCTCCATATAATAACATCAGAAAAAGAT 2424

Qy       607 ArgIleAlaProValHisIleAspThrGluSer---IleSerAlaLeuIleLysGlnVal 625
             ::::::            :::      ||||||   :::      |||::::::::::::
Db      2425 AAATTGTTTTCTGAAGTAGTTCATAAGGAGAGTAGAGTTCAAGGTTTACTTGAAGAAATT 2484

Qy       626 AsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAla 645
             |||:::      |||      |||      :::               |||||||||
Db      2485 GGGAAAACAAAAGATGACCTAGCAACTACACAGTCGAATTATAAAAGCACTGATCAAGAA 2544

Qy       646 Ile--------------------------------------------------------- 646

Db      2545 TTCCAAAATTTCAAAACCCTTCATATGGACTTTGAGCAAAAGTATAAGATGGTCCTTGAG 2604

Qy       647 ---------ArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSer 663
                      |||         |||::::::   :::
Db      2605 GAGAATGAGAGAATGAATCAGGAAATAGTTAATCTCTCTAAA------------------ 2646

Qy       664 PheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp--------- 680
                      |||:::   |||:::|||   |||      ||||||   :::
Db      2647 ------GAAGCCCAAAAATTTGATTCGAGTTTGGGTGCTTTGAAGACCGAGCTTTCTTAC 2700

Qy       681 ---------------AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThr 695
                               ||||||      |||         ::::::      |||
Db      2701 AAGACCCAAGAACTTCAGGAGAAAACACGTGAGGTTCAAGAAAGACTA------AATGAG 2754

Qy       696 GlySerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHis 715
                   :::::::::|||         :::      |||   |||      |||
Db      2755 ATGGAACAGCTGAAGGAACAATTAGAAAATAGAGATTCTCCGCTGCAAACTGTAGAAAGG 2814

Qy       716 HisLysSer--------------------------------------------------- 718
             |||:::
Db      2815 GAGAAAACACTGATTACTGAGAAACTGCAGCAAACTTTAGAAGAAGTAAAAACTTTAACT 2874

Qy       719 ---LysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePhe 737
                :::|||            :::|||   :::   ::::::
Db      2875 CAAGAAAAAGATGATCTAAAACAACTCCAAGAAAGCTTGCAAATT--------------- 2919

Qy       738 SerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro 757
                   ::::::         :::   |||   :::   ::::::   ::::::|||
Db      2920 ---GAGAGGGACCAACTCAAAAGTGATATTCACGATACTGTTAACATGAATATAGATACT 2976

Qy       758 SerAsn---------------LeuGluHisLeuIleThrProLeuValThrIleGlyHis 772
                                  ||||||   |||               ||||||
Db      2977 CAAGAACAATTACGAAATGCTCTTGAGTCTCTGAAACAACATCAAGAAACAATTAATACA 3036

Qy       773 IleAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPhe 792
             :::                                    ||||||   ::::::
Db      3037 CTA------------------------------------AAATCGAAAATTTCTGAGGAA 3060

Qy       793 IleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrp 812
             :::   ::::::|||   |||
Db      3061 GTTTCCAGGAATTTGCATATG--------------------------------------- 3081

Qy       813 ValProAspGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMet 832
                      ||||||   :::   ||||||      :::   |||
Db      3082 ---------GAGGAAAATACAGGAGAAACTAAAGATGAATTTCAG--------------- 3117

Qy       833 ValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg 852
                :::   ::::::|||:::                           :::|||   :::
Db      3118 ---CAAAAGATGGTTGGCATAGATAAAAAACAGGATTTGGAAGCTAAAAATACC---CAA 3171

Qy       853 LeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLys 872
                ||||||   :::      :::      :::   :::|||      ||||||
Db      3172 ACACTAACTGCAGATGTTAAGGATAATGAGATAATTGAGCAACAAAGGAAGATATTTTCT 3231

Qy       873 ------ProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAla 890
                :::   :::   |||:::                  |||   |||   :::
Db      3232 TTAATACAGGAGAAAAATGAACTCCAACAAATGTTAGAGAGTGTTATAGCAGAAAAGGAA 3291

Qy       891 Gln------------------GluProCysTyrHisGluIleIleThrLeuGluGlnTyr 904
             |||                  |||               |||         |||:::
Db      3292 CAATTGAAGACTGACCTAAAGGAAAATATTGAAATGACCATTGAAAACCAGGAAGAATTA 3351

Qy       905 GlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGln 924
             :::|||            :::   ||||||      :::   :::::::::   ||||||
Db      3352 AGACTT------------CTTGGGGATGAACTTAAAAAGCAACAAGAGATAGTTGCACAA 3399

Qy       925 LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAla 944
             :::   :::      :::   :::               |||            |||
Db      3400 GAAAAGAACCATGCCATAAAGAAA---------GAAGGAGAGCTTTCTAGGACCTGTGAC 3450

Qy       945 LeuCysAla------LysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeu 962
                   |||      ::::::   :::||||||:::         ::::::
Db      3451 AGACTGGCAGAAGTTGAAGAAAAACTAAAGGAAAAGAGCCAGCAACTCCAAGAAAAACAG 3510

Qy       963 ValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGlu--- 981
                :::   :::||||||:::   |||      :::            ::::::|||
Db      3511 CAACAACTTCTTAATGTACAAGAAGAGATGAGTGAGATGCAGAAAAAGATTAATGAAATA 3570

Qy       982 ---------------------LysLeuLeuSerLeuLeuProGluTyrValValProTyr 994
                                  |||         |||      |||::::::
Db      3571 GAGAATTTAAAGAATGAATTAAAGAACAAAGAATTGACATTGGAACATATGGAAACAGAG 3630

Qy       995 ThrIleHisLeuLeuAlaHis---AspProAspTyrValLysValGlnAspIle---Glu 1012
                :::   |||            :::   :::|||   :::|||:::   |||   :::
```

```
Db      3631 AGGCTTGAGTTGGCTCAGAAACTTAATGAAAATTATGAGGAAGTGAAATCTATAACCAAA 3690

Qy      1013 GlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsn 1032
             :::  |||  :::||||||              |||:::
Db      3691 GAAAGAAAAGTTCTAAAGGAA--------------TTACAGAAGTCATTTGAAACAGAG 3735

Qy      1033 GluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAsp 1052
                ::::::          :::||||||::::::                |||||||||:::
Db      3736 AGAGACCACCTTAGAGGATATATAAGAGAAATTGAAGCTACAGGCCTACAAACCAAAGAA 3795

Qy      1053 AlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAla 1072
                                      :::   |||       |||:::   ::::::
Db      3796 GAACTAAAAATTGCTCATATTCACCTAAAAGAACACCAAGAAACTATTGATGAACTAAGA 3855

Qy      1073 MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProVal 1092
                ::::::     |||:::   |||
Db      3856 AGAAGCGTA------TCTGAGAAGACA-------------------------------- 3876

Qy      1093 LeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeu 1112
                   |||:::               |||        |||:::||||||       |||
Db      3877 ------GCTCAAATAATAAATACTCAGGACTTAGAAAAATCCCATACCAAA------TTA 3924

Qy      1113 ProProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGly 1132
                   |||:::                                        ||||||
Db      3925 CAAGAAGAGATCCCA-----------------------------------GTGCTTCAT 3948

Qy      1133 AlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSer------Ser 1150
                 :::   |||   :::           |||   |||:::|||:::
Db      3949 GAGGAACAAGAGTTACTGCCTAATGTGAAAAAAGTCAGTGAGACTCAGGAAACAATGAAT 4008

Qy      1151 ArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIle 1170
                :::|||   ::::::      |||::::::   :::   ::::::        ||||||
Db      4009 GAACTGGAGTTATTAACAGAACAGTCCACAACCAAGGACTCAACAACACTGGCAAGAATA 4068

Qy      1171 LysGlyArgLeuAspSerSerGluMetAspHisSerGlu---AsnGluAspTyrThrMet 1189
                                       ||||||:::         ||||||   :::
Db      4069 --------------------GAAATGGAAAGGCTCAGGTTGAATGAAAAATTTCAAGAA 4107

Qy      1190 SerSerProLeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgSer 1209
             |||                      :::   :::||||||:::            ::::::
Db      4108 AGTCAGGAAGAGATAAAATCTCTAACCAAGGAAAGAGACAACCTTAAAACGATAAAAGAA 4167

Qy      1210 GluLeuGluLysProArgGlyArgLysLysThrProValThrGlu--------------- 1224
                ||||||              :::   |||       :::   |||
Db      4168 GCCCTTGAAGTTAAACATGACCAGCTGAAAGAACATATTAGAGAAACTTTGGCTAAAATC 4227

Qy      1225 -----------------GlnGluGluLysLeuGlyMet-----------AspAspLeu 1234
                                 ||||||:::   |||   |||            ::::::
Db      4228 CAGGAGTCTCAAAGCAAACAAGAACAGTCCTTAAATATGAAAGAAAAAGACAATGAAACT 4287

Qy      1235 ThrLysLeuValGlnGluGln---------LysProLys--------------------- 1244
             ||||||:::|||   |||         |||||||||
Db      4288 ACCAAAATCGTGAGTGAGATGGAGCAATTCAAACCCAAAGATTCAGCACTACTAAGGATA 4347

Qy      1245 --------------GlySerGlnArgSerArgLysArgGlyHis--------------- 1254
                           |||   :::      :::   |||
Db      4348 GAAATAGAAATGCTCGGATTGTCCAAAAGACTTCAAGAAAGTCATGATGAAATGAAATCT 4407

Qy      1255 ThrAlaSerGluSerAspGluGlnGln------------------TrpProGluGluLys 1268
                |||   |||   |||:::   |||                          |||
Db      4408 GTAGCTAAGGAGAAAGATGACCTACAGAGGCTGCAAGAAGTTCTTCAATCTGAAAGTGAC 4467

Qy      1269 ArgLeuLysGluAspIle--------------------LeuGluAsnGluAsp------ 1279
             :::||||||||||:::|||                      ||||||   |||:::
Db      4468 CAGCTCAAAGAAAACATAAAAGAAATTGTAGCTAAACACCTGGAAACTGAAGAGGAACTT 4527

Qy      1280 -----------------------GluGlnAsnSerProProLysLysGlyLysArgGly 1291
                                     ||||||            :::   :::
Db      4528 AAAGTTGCTCATTGTTGCCTGAAAGAACAAGAGGAAACTATTAATGAGTTAAGAGTGAAT 4587

Qy      1292 ArgProProLysProLeuGlyGlyGlyThrProLysGluGluProThrMetLysThrSer 1311
                      |||                          |||       |||::::::
Db      4588 CTTTCAGAGAAG------------------GAAACTGAAATATCAACCATTCAAAAGCAG 4629

Qy      1312 LysLysGlySerLysLysLysSerGlyProProAlaProGluGluGluGluGluGluGlu 1331
                :::               |||                |||       |||:::||||||
Db      4630 TTAGAAGCAATCAATGATAAATTACAGAACAAGATCCAAGAGATTTATGAGAAAGAGGAA 4689

Qy      1332 ------------ArgGlnSerGlyAsnThrGluGlnLys---SerLysSerLysGlnHis 1346
                         :::|||           :::::::|||   ::::::   ||||||
Db      4690 CAACTTAATATAAAACAAATTAGTGAGGTTCAGGAAAAAGTGAATGAACTGAAACAATTC 4749

Qy      1347 ArgValSerArgArgAlaGlnGlnArgAla---GluSerProGluSerSerAlaIleGlu 1365
                :::      |||:::|||:::          |||   :::|||   ||||||     :::|||
Db      4750 AAGGAGCATCGCAAAGCCAAGGATTCAGCACTACAAAGTATAGAAAGTAAGATGCTCGAG 4809

Qy      1366 SerThrGlnSerThrProGlnLysGlyArg---------------------------- 1375
                |||   |||   :::|||   |||
Db      4810 TT-GACCAACAGACTTCAAGAAAGTCAAGAAGAAATACAAATTATGATTAAGGAAAAAGA 4868

Qy      1376 -----------------GlyArgProSerLysThrProSerProSerGlnProLys 1388
                              |||   ||||||            |||::::::   |||
Db      4869 GGAAATGAAAAGAGTACAGGAGGCCCTTCAGATAGAGAGAGACCAACTGAAAGAAAA 4925
```

RESULT 3
US-09-595-684B-30
; Sequence 30, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: VAisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; TITLE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30
;   LENGTH: 8257
;   TYPE: DNA
;   ORGANISM: Human
US-09-595-684B-30

Alignment Scores:

Pred. No.:	9.88e-11	Length:	8257
Score:	229.00	Matches:	285
Percent Similarity:	35.61%	Conservative:	268
Best Local Similarity:	18.35%	Mismatches:	546
Query Match:	3.18%	Indels:	455
DB:	4	Gaps:	66

US-09-512-581B-2 (1-1391) x US-09-595-684B-30 (1-8257)

```
Qy        17 ProGlyValLysGluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLys 36
             |||   |||   |||:::|||               :::   |||   :::::::::|||   :::
Db      1087 CCTTATGTTAATGAGGTATCA---------ACTGATGAAGCTCTCCTGAAAAGGTATAGA 1137

Qy        37 MetValValLysThrPheMetAspMetAspGlnAspSerGluGlu-------------- 51
                      |||   |||   :::|||   :::      ||||||
Db      1138 ---------AAAGAAATAATGGATCTTAAAAAACAATTAGAGGAGGTTTCTTTAGAGACG 1188

Qy        52 ---------------GluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
                            ||||||:::                |||                |||
Db      1189 CGGGCTCAGGCAATGGAAAAAGACCAATTGGCCCAACTTTTGGAAGAA------AAAGAT 1242

Qy        67 PhePheLeuLysHisProGlyLysAspVal-----------ArgLeuLeuValAlaCys 82
                      |||      :::   :::                  |||:::||||||
Db      1243 TTGCTTCAGAAAGTACAGAATGAGAAAATTGAAAACTTAACACGGATGCTGGTG------ 1296
```

```
Qy          83 CysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLys 102
Db        1297 -------------------------------------------ACCTCTTCTTCC--- 1308

Qy         103 LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLysSer 122
Db        1309 --------------------CTCACGTTGCAA-----------CAGGAATTAAAGGCT 1335

Qy         123 ProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsn 142
Db        1336 AAAGAAAACGAAGAGTTACTTGGTGCCTTGGCAAAATTAACAAAATGAAGAACTCAAAC 1395

Qy         143 IleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPhe 162
Db        1396 -------------------------------------------TATGCAGATCAATTT 1410

Qy         163 SerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSer 182
Db        1411 AATATACCAACAAAT---ATAACAACAAAAACACATAAGCTTTCTATAAATTTATTACGA 1467

Qy         183 SerIle---------IleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrVal 199
Db        1468 GAAATTGATGAATCTGTCTGTTCAGAGTCTGATGTTTTCAGTAACACTCTTGATACATTA 1527

Qy         200 Leu---ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAla 218
Db        1528 AGTGAGATAGAATGGAATCCAGCAACAAAGCTACTAAATCAGGAGAATATAGAA------ 1581

Qy         219 LysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsn 238
Db        1581 ------------------------------------------------------------ 1581

Qy         239 GlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeu 258
Db        1582 --------------------AGTGAGTTGAACTCACTTCGTGCTGACTATGATAATCTG 1620

Qy         259 IleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGlu 278
Db        1621 GTATTA---------------GACTATGAACAACTACGAACAGAAAAAGAAGAAATGGAA 1665

Qy         279 PheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLys 298
Db        1666 TTGAAATTAAAAGAAAAGAATGATTTGGATGAATTTGAGGCTCTAGAAAGAAAAACTAAA 1725

Qy         299 MetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyr 318
Db        1726 ------------AAAGATCAAGAGATGCAA---------------------------- 1743

Qy         319 LeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSer 338
Db        1743 ------------------------------------------------------------ 1743

Qy         339 HisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArg 358
Db        1744 ------CTAATTCATGAAATTTCGAACTTAAAGAATTTAGTTAAGCATCGAGAAGTATAT 1797

Qy         359 SerHisAspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAla 378
Db        1798 AATCAAGATCTTGAGAATGAACTC-----------------------------AGTTCA 1827

Qy         379 LysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeu 398
Db        1828 AAAGTAGAGCTGCTTAGAGAAAAGGAAGACCAGATTAAGAAGCTACAGGAATACATAGAC 1887

Qy         399 AspLysArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLys 418
Db        1888 TCTCAAAAG---------CTAGAAAATATAAAAATGGACTTGTCA-------------- 1923

Qy         419 TyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAsp 438
Db        1924 TACTCATTGGAAAGCATTGAAGACCCAAAACAAATGAAGCAGACTCTGTTTGATGCTGAA 1983

Qy         439 LysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIle 458
Db        1984 ACTGTA-----------------------------------GCCCTTGATGCCAAGAGA 2007

Qy         459 PheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyr 478
Db        2008 GAATCAGCCTTTCTTAGAAGTGAAAATCTGGAGTTGAAGGAGAAAATGAAAGAACTTGCA 2067

Qy         479 TyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCys 498
Db        2068 ACTACATACAAGCAAATGGAA-------------------AATGATATTCAGTTATAT 2106

Qy         499 GlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThr 518
Db        2107 CAAAGCCAATTGGAGGCAAAAAAGAAAATGCAAGTTGATCTGGAGAAAGAA--------- 2157

Qy         519 AspAlaSerValLysAlaIlePheSerLysValMetValIleThrArg----------- 534
Db        2158 ---------TTACAATCTGCTTTTAATGAGATAACAAAACTCACCTCCCTTATAGATGGC 2208

Qy         535 -----------------------AsnLeuProAspProGlyLysAlaGlnAspPheMet 546
Db        2209 AAAGTTCCAAAAGATTTGCTCTGTAATTTGGAATTGGAAGGAAAGATTACTGATCTTCAG 2268

Qy         547 LysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeu 566
Db        2269 AAAGAACTAAATAAAGAAGTTGAAGAAAATGAAGCTTTGCGGGAAGAAGTC---ATTTTG 2325

Qy         567 ValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLys 586
Db        2326 CTTTCA---------GAATTGAAATCTTTACCTTCTGAAGTAGAAAGGCTGAGGAAAGAG 2376

Qy         587 LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGlu 606
Db        2377 ATA------------CAAGACAAATCTGAAGAGCTCCATATAATAACATCAGAAAAAGAT 2424

Qy         607 ArgIleAlaProValHisIleAspThrGluSer---IleSerAlaLeuIleLysGlnVal 625
Db        2425 AAATTGTTTTCTGAAGTAGTTCATAAGGAGAGTAGAGTTCAAGGTTTACTTGAAGAAATT 2484

Qy         626 AsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAla 645
Db        2485 GGGAAAACAAAAGATGACCTAGCAACTACACAGTCGAATTATAAAAGCACTGATCAAGAA 2544

Qy         646 Ile--------------------------------------------------------- 646
Db        2545 TTCCAAAATTTCAAAACCCTTCATATGGACTTTGAGCAAAAGTATAAGATGGTCCTTGAG 2604

Qy         647 ---------ArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSer 663
Db        2605 GAGAATGAGAGAATGAATCAGGAAATAGTTAATCTCTCTAAA----------------- 2646

Qy         664 PheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp--------- 680
Db        2647 ------GAAGCCCAAAAAATTTGATTCGAGTTTGGGTGCTTTGAAGACCGAGCTTTCTTAC 2700

Qy         681 ---------------AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThr 695
Db        2701 AAGACCCAAGAACTTCAGGAGAAAACACGTGAGGTTCAAGAAAGACTA------AATGAG 2754

Qy         696 GlySerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHis 715
Db        2755 ATGGAACAGCTGAAGGAACAATTAGAAAATAGAGATTCTCCGCTGCAAACTGTAGAAAGG 2814

Qy         716 HisLysSer--------------------------------------------------- 718
Db        2815 GAGAAAACACTGATTACTGAGAAACTGCAGCAAACTTTAGAAGAAGTAAAAACTTTAACT 2874

Qy         719 ---LysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePhe 737
Db        2875 CAAGAAAAAGATGATCTAAAACAACTCCAAGAAAGCTTGCAAATT-------------- 2919

Qy         738 SerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro 757
```

```
Db   2920 ---GAGAGGGACCAACTCAAAAGTGATATTCACGATACTGTTAACATGAATATAGATACT 2976

Qy    758 SerAsn---------------LeuGluHisLeuIleThrProLeuValThrIleGlyHis 772
                               ||||||  |||                 ||||||
Db   2977 CAAGAACAATTACGAAATGCTCTTGAGTCTCTGAAACAACATCAAGAAACAATTAATACA 3036

Qy    773 IleAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPhe 792
          :::                                      ||||||  ::::::
Db   3037 CTA-----------------------------------AAATCGAAAATTTCTGAGGAA 3060

Qy    793 IleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrp 812
          :::   ::::::|||   |||
Db   3061 GTTTCCAGGAATTTGCATATG--------------------------------------- 3081

Qy    813 ValProAspGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMet 832
                   ||||||   :::   ||||||      :::   |||
Db   3082 ---------GAGGAAAATACAGGAGAAACTAAAGATGAATTTCAG--------------- 3117

Qy    833 ValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArg 852
             :::   ::::::|||:::                           :::|||   :::
Db   3118 ---CAAAAGATGGTTGGCATAGATAAAAAACAGGATTTGGAAGCTAAAAATACC---CAA 3171

Qy    853 LeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLys 872
             ||||||      :::      :::      :::   :::|||   ||||||
Db   3172 ACACTAACTGCAGATGTTAAGGATAATGAGATAATTGAGCAACAAAGGAAGATATTTTCT 3231

Qy    873 ------ProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAla 890
                :::   :::   |||:::               |||   |||   :::
Db   3232 TTAATACAGGAGAAAAATGAACTCCAACAAATGTTAGAGAGTGTTATAGCAGAAAAGGAA 3291

Qy    891 Gln------------------GluProCysTyrHisGluIleIleThrLeuGluGlnTyr 904
          |||                  |||               |||      |||:::
Db   3292 CAATTGAAGACTGACCTAAAGGAAAATATTGAAATGACCATTGAAAACCAGGAAGAATTA 3351

Qy    905 GlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGln 924
          :::|||            :::   ||||||      :::   :::::::::   ||||||
Db   3352 AGACTT------------CTTGGGGATGAACTTAAAAAGCAACAAGAGATAGTTGCACAA 3399

Qy    925 LysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAla 944
          :::   :::      :::   :::                  |||            |||
Db   3400 GAAAAGAACCATGCCATAAAGAAA---------GAAGGAGAGCTTTCTAGGACCTGTGAC 3450

Qy    945 LeuCysAla------LysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeu 962
             |||         ::::::   :::|||||||:::         ::::::
Db   3451 AGACTGGCAGAAGTTGAAGAAAAACTAAAGGAAAAGAGCCAGCAACTCCAAGAAAAACAG 3510

Qy    963 ValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGlu--- 981
             :::   :::|||||||:::   |||         :::         ::::::|||
Db   3511 CAACAACTTCTTAATGTACAAGAAGAGATGAGTGAGATGCAGAAAAAGATTAATGAAATA 3570

Qy    982 --------------------LysLeuLeuSerLeuLeuProGluTyrValValProTyr 994
                              |||         |||      |||::::::
Db   3571 GAGAATTTAAAGAATGAATTAAAGAACAAAGAATTGACATTGGAACATATGGAAACAGAG 3630

Qy    995 ThrIleHisLeuLeuAlaHis---AspProAspTyrValLysValGlnAspIle---Glu 1012
          :::   |||                 :::   :::|||   :::|||:::   |||   :::
Db   3631 AGGCTTGAGTTGGCTCAGAAACTTAATGAAAATTATGAGGAAGTGAAATCTATAACCAAA 3690

Qy   1013 GlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsn 1032
          :::   |||   :::||||||                  |||:::
Db   3691 GAAAGAAAAGTTCTAAAGGAA---------------TTACAGAAGTCATTTGAAACAGAG 3735

Qy   1033 GluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAsp 1052
             ::::::         :::||||||::::::               |||||||||:::
Db   3736 AGAGACCACCTTAGAGGATATATAAGAGAAATTGAAGCTACAGGCCTACAAACCAAAGAA 3795

Qy   1053 AlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAla 1072
                                 :::   |||         |||:::   ::::::
Db   3796 GAACTAAAAATTGCTCATATTCACCTAAAAGAACACCAAGAAACTATTGATGAACTAAGA 3855

Qy   1073 MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProVal 1092
             ::::::      |||:::   |||
Db   3856 AGAAGCGTA------TCTGAGAAGACA--------------------------------- 3876

Qy   1093 LeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeu 1112
                |||:::               |||      |||:::||||||         |||
Db   3877 ------GCTCAAATAATAAATACTCAGGACTTAGAAAAATCCCATACCAAA------TTA 3924

Qy   1113 ProProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGly 1132
                |||:::                                          ||||||
Db   3925 CAAGAAGAGATCCCA----------------------------------GTGCTTCAT 3948

Qy   1133 AlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSer------Ser 1150
                :::   |||   :::         |||   |||:::|||:::            :::
Db   3949 GAGGAACAAGAGTTACTGCCTAATGTGAAAAAAGTCAGTGAGACTCAGGAAACAATGAAT 4008

Qy   1151 ArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIle 1170
             :::|||   ::::::      |||:::::::   :::   ::::::      ||||||
Db   4009 GAACTGGAGTTATTAACAGAACAGTCCACAACCAAGGACTCAACAACACTGGCAAGAATA 4068

Qy   1171 LysGlyArgLeuAspSerSerGluMetAspHisSerGlu---AsnGluAspTyrThrMet 1189
                                 ||||||:::         ||||||   :::
Db   4069 --------------------GAAATGGAAAGGCTCAGGTTGAATGAAAAATTTCAAGAA 4107

Qy   1190 SerSerProLeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgSer 1209
          |||                     :::   :::||||||:::         ::::::
Db   4108 AGTCAGGAAGAGATAAAATCTCTAACCAAGGAAAGAGACAACCTTAAAACGATAAAAGAA 4167

Qy   1210 GluLeuGluLysProArgGlyArgLysLysThrProValThrGlu--------------- 1224
             ||||||               :::   |||      :::   |||
Db   4168 GCCCTTGAAGTTAAACATGACCAGCTGAAAGAACATATTAGAGAAACTTTGGCTAAAATC 4227

Qy   1225 ------------------GlnGluGluLysLeuGlyMet------------AspAspLeu 1234
                            ||||||:::   |||   |||            ::::::
Db   4228 CAGGAGTCTCAAAGCAAACAAGAACAGTCCTTAAATATGAAAGAAAAAGACAATGAAACT 4287

Qy   1235 ThrLysLeuValGlnGluGln---------LysProLys--------------------- 1244
          ||||||:::|||   |||            |||||||||
Db   4288 ACCAAAATCGTGAGTGAGATGGAGCAATTCAAACCCAAAGATTCAGCACTACTAAGGATA 4347

Qy   1245 ---------------GlySerGlnArgSerArgLysArgGlyHis--------------- 1254
                         |||      :::      :::      |||
Db   4348 GAAATAGAAATGCTCGGATTGTCCAAAAGACTTCAAGAAAGTCATGATGAAATGAAATCT 4407

Qy   1255 ThrAlaSerGluSerAspGluGlnGln------------------TrpProGluGluLys 1268
             |||   |||   |||:::   |||                        |||
Db   4408 GTAGCTAAGGAGAAAGATGACCTACAGAGGCTGCAAGAAGTTCTTCAATCTGAAAGTGAC 4467

Qy   1269 ArgLeuLysGluAspIle---------------------LeuGluAsnGluAspGluGln 1281
          :::||||||||||||:::|||                  ||||||   |||:::|||
Db   4468 CAGCTCAAAGAAAACATAAAAGAAATTGTAGCTAAACACCTGGAAACTGAAGAGGAACTT 4527

Qy   1282 AsnSerProProLysLysGlyLysArgGlyArgProPro---LysProLeuGlyGlyGly 1300
                               |||                     |||
Db   4528 AAAGTTGCTCATTGTTGCCTGAAAGAACAAGAGGAAACTATTAATGAGTTAAGAGTGAAT 4587

Qy   1301 ThrProLysGluGluProThrMetLysThrSerLysLys---------------GlySer 1315
                :::::::|||      :::   |||   :::|||
Db   4588 CTTTCAGAGAAGGAAACTGAAATATCAACCATTCAAAAGCAGTTAGAAGCAATCAATGAT 4647

Qy   1316 LysLysLysSerGlyProProAlaProGluGluGluGluGlu---------GluArg 1332
          |||      ::::::                  |||:::||||||:::         ::::::
Db   4648 AAATTACAGAACAAGATCCAAGAGATTTATGAGAAAGAGGAACAACTTAATATAAAACAA 4707

Qy   1333 GlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgValSerArgArgAla 1352
             |||                  |||   ::::::   ||||||   :::      |||:::|||
Db   4708 ATTAGTGAGGTTCAGGAAAACGTGAATGAACTGAAACAATTCAAGGAGCATCGCAAAGCC 4767

Qy   1353 GlnGlnArgAla---GluSerProGluSerSerAlaIleGluSerThrGlnSerThrPro 1371
          :::         |||   :::|||   ||||||      :::|||      |||   |||
Db   4768 AAGGATTCAGCACTACAAAGTATAGAAAGTAAGATGCTCGAGTT-GACCAACAGACTTCA 4826

Qy   1372 GlnLysGlyArg------------------------------------------------ 1375
          :::|||   |||
Db   4827 AGAAAGTCAAGAAGAAATACAAATTATGATTAAGGAAAAAGAGGAAATGAAAAGAGTACA 4886
```

```
Qy       1376 GlyArgProSerLysThrProSerProSerGlnProLys 1388
               |||   ||||||              |||:::::   |||
Db       4887 GGAGGCCCTTCAGATAGAGAGAGACCAACTGAAAGAAAA 4925

RESULT 4
US-08-931-999-4
; Sequence 4, Application US/08931999
; Patent No. 6043219
;  GENERAL INFORMATION:
;    APPLICANT:  Iandolo, John J.
;    APPLICANT:  Crupper, Scott S.
;    TITLE OF INVENTION:  Broad Spectrum Chemotherapeutic Peptide
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hovey, Williams, Timmons & Collins
;      STREET:  2405 Grand Boulevard, Suite 400
;      CITY:  Kansas City
;      STATE:  Missouri
;      COUNTRY:  U.S.A.
;      ZIP:  64108
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/931,999
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/710,561
;      FILING DATE:  19-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Collins, John M.
;      REGISTRATION NUMBER:  26,262
;      REFERENCE/DOCKET NUMBER:  25043-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  816/474-9050
;      TELEFAX:  816/474-9057
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6755 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Staphylococcus aureus
;      STRAIN:  UT0007
US-08-931-999-4

Alignment Scores:
Pred. No.:              9.22e-09      Length:        6755
Score:                  207.00        Matches:       78
Percent Similarity:     42.81%        Conservative:  50
Best Local Similarity:  26.09%        Mismatches:    137
Query Match:            2.88%         Indels:        36
DB:                     3             Gaps:          9

US-09-512-581B-2 (1-1391) x US-08-931-999-4 (1-6755)

Qy       1121 ThrProGlyLysProLys---ThrThrAsnValLeuGlyAlaValAsnLysProLeuSer 1139
               |||    |||||||||  ||||||                          :::|||
Db       5664 ACAGCACAAAAACCCAAGAACACGACAAAAGGAGAAAAAGAAAAAAGACAACCAAGACAA 5723

Qy       1140 SerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSer 1159
                        ||||||  ||||||:::      :::  :::
Db       5724 GAAAAAAGAAAACAAAGACAAACAAGAAAAGAAAAGAAGAAAAAAAACCAGCGACCCCGA 5783

Qy       1160 SerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMet 1179
                                       |||:::  :::|||:::           |||
Db       5784 CAGAAAAGAAGAGCAGAGGGAAACGGACAACAACAGGGACAGGGACAACACCCAGAAACA 5843

Qy       1180 AspHisSerGluAsnGluAspTyrThrMetSerSerPro---LeuProGlyLysLysSer 1198
                    :::      ::::::                           ||||||:::
Db       5844 CCAACAACACGAACCAAAGAGAAAGCAAGAGAGCAAAAAAACACAACAGGGAAGAGGAAA 5903

Qy       1199 AspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLys 1218
                 |||   ::::::             :::::::::  |||||||||||       |||
Db       5904 ACCAAAGAGGAAGAGCAAAAAAG-AAAAAGGCACAGCCAGAGAAACCAAGAAAGGGAAAA 5962

Qy       1219 Lys-ThrProValThrGluGlnGluGlu---------LysLeuGlyMetAspAspLeuTh 1235
              ||| |||           :::|||||||||         |||   |||        :::
Db       5963 AAACACGAACGGAAAAAAACAAGAGGAAAAGAAGAACAAAAAAGGAGAACGAGAAAAACA 6022

Qy       1235 rLysLeuValGlnGluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisTh 1255
               :::         ||||||:::    :::       ::::::   |||:::   |||
Db       6023 AAGAAACAGGAACGAGCAAGAAGAAGAACAACAAAAAAAAAGAAGGAGGGAAGGAAGAAA 6082

Qy       1255 rAlaSerGluSerAspGluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLe 1275
                       |||   |||:::::::::   |||   ||||||         :::
Db       6083 GGAGCAGGAAAA-GACCAAGAAAAAACACCAACCGAAAAAGCACGGAGA----------- 6130

Qy       1275 uGluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgProProLy 1295
               |||         ::::::         |||||||||   |||   |||:::      ||
Db       6131 -GAAGGAACGAAGAAGGAAAAGAAGAACCCAAAAAAAAACAAAACGGGACAGAGCACAAA 6189

Qy       1295 sProLeuGlyGlyGlyThrProLysGluGluProThrMetLysThrSerLysLysGlySe 1315
               |                      ::::::|||         |||      |||:::
Db       6190 GACCAAAAACAACAACGAAAAACGACAAGAAACAAAACACAAAGAGAGAAAAGAAAAGA 6249

Qy       1315 rLysLys------------------------------------LysSerGlyProPr 1322
               ||||||                                      :::     ||
Db       6250 AAAAAAACGCAGGAGAAACAAAAGCAGGGAAACAAGCAAAAAAACCCGCAAAGACGAGCC 6309

Qy       1322 oAlaProGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys----- 1340
               |     |||||||||       ::::::|||:::      :::   :::||||||
Db       6310 CACAAAAGAGGAAGAAACCGCGAAAAAAAGAAAAAAAAAAAGCCAAAAACAAAAAGGAAC 6369

Qy       1341 ----SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerPr 1359
                  ||||||  |||:::   |||       :::|||    :::||||||    :::   ||
Db       6370 AACAAGCAAAAAGAAGAAGAAAAGGGAAGAAAAAAGAGGAAAACAGAGAGAACAAAAGCC 6429

Qy       1359 o------GluSerSerAlaIleGluSerThrGlnSerThrProGlnLysGly---ArgGl 1376
               |      :::   |||      |||      ::::::      |||||||||    :::
Db       6430 AAAAAAACAACGAAGCGGAGAAGAAAGAGCAAAAACAAACACGCAAAAGGGCAAAAAAGC 6489

Qy       1376 yArgProSerLys---------------ThrProSerProSerGlnProLysLys 1389
               ||||||:::|||                 |||         :::||||||||||||
Db       6490 ACGACCAAACAAAAGACAAACAAGCACAACAAAGCGAACAACACAACCCAAAAAA 6544

RESULT 5
US-08-353-700-2
; Sequence 2, Application US/08353700
; Patent No. 5599919
;  GENERAL INFORMATION:
;    APPLICANT:  YEN, TIMOTHY J.
;    APPLICANT:  RATTNER, JEROME B.
;    TITLE OF INVENTION:  NUCLEIC ACID ENCODING A
;    TITLE OF INVENTION:  TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DANN, DORFMAN, HERRELL AND SKILLMAN
;      STREET:  1601 MARKET STREET, SUITE 720
;      CITY:  PHILADELPHIA
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2307
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/353,700
;      FILING DATE:  09-DEC-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  REED, JANET E.
;      REGISTRATION NUMBER:  36,252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-4100
;      TELEFAX:  (215) 563-4044
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10136 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  HUMAN
US-08-353-700-2

Alignment Scores:
Pred. No.:              6.44e-08        Length:          10136
Score:                  201.50          Matches:         283
Percent Similarity:     35.16%          Conservative:    242
Best Local Similarity:  18.96%          Mismatches:      587
Query Match:            2.80%           Indels:          382
DB:                     1               Gaps:            63

US-09-512-581B-2 (1-1391) x US-08-353-700-2 (1-10136)

Qy         19 ValLysGluIleSerAspLysIle-----------------------SerLysGluGlu 30
              :::|||||:::      |||:::                         :::|||   |||
Db       5868 ATGAAAGAATTAGACTCAAAACTCCATTTACAGGAGGTACAACTAATGACCAAAATTGAA 5927

Qy         31 MetValArgArgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGlu 50
                                |||:::   :::|||               :::    |||
Db       5928 GCATGCATAGAATTGGAAAAAATAGTTGGGGAACTTAAGAAAGAAAACTCAGATTTAAGT 5987

Qy         51 GluGluLysGluLeuTyr------------------------------------------ 56
              |||:::   |||   :::
Db       5988 GAAAAATTGGAATATTTTTCTTGTGATCACCAGGAGTTACTCCAGAGAGTAGAAACTTCT 6047

Qy         57 ------LeuAsnLeuAlaLeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLys 74
                    ||||||      |||    :::   :::|||         |||
Db       6048 GAAGGCCTCAATTCTGATTTAGAAATGCATGCAGAT---------AAATCATCACGTGAA 6098

Qy         75 AspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGlu 94
              |||:::            ||||||
Db       6099 GATATTGGAGATAATGTGGCC--------------------------------------- 6119

Qy         95 AlaProTyrThrSerProAspLysLeuLysAspIlePheMetPheIleThrArgGlnLeu 114
                                 |||       |||:::   |||:::   :::      :::|||
Db       6120 ---------AAGGTGAATGACAGCTGGAAGGAGAGATTTCTTGATGTGGAAAATGAGCTG 6170

Qy        115 LysGlyLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsn 134
                    :::             |||:::         :::         ||||||      :::
Db       6171 AGTAGGATCAGATCGGAGAAAGCTAGCATTGAGCATGAAGCCCTCTACCTGGAGGCTGAC 6230

Qy        135 IleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSerAsn------Glu 152
              :::      |||::::::         :::|||      ||||   ::::::   |||          :::
Db       6231 TTAGAGGTAGTTCAAACAGAGAAGCTATGTTTAGAAAAAGACAATGAAAATAAGCAGAAG 6290

Qy        153 IlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAsnGlyHisAsn------ 170
              :::         |||         |||   ||||||:::   :::         |||
Db       6291 GTTATTGTCTGCCTTGAAGAAGAACTC---TCAGTGGTCACAAGTGAGAGAAACCAGCTT 6347

Qy        170 ------------------------------------------------------------ 170

Db       6348 CGTGGAGAATTAGATACTATGTCAAAAAAAACCACGGCACTGGATCAGTTGTCTGAAAAA 6407
```

```
Qy        171 --------------GlnLysValHisMetHisMetValAspLeuMetSerSerIle--- 184
                            |||::::::         |||         :::   :::   |||
Db       6408 ATGAAGGAGAAAACACAAGAGCTTGAGTCTCATCAAAGTGAGTGTCTCCATTGCATTCAG 6467

Qy        185 IleCysGluGlyAspThrValSerGln---GluLeuLeuAspThrValLeuValAsnLeu 203
              :::   |||   :::            :::   ||||||||   |||:::            ::::::
Db       6468 GTGGCAGAGGCAGAGGTGAAGGAAAAGACGGAACTCCTTCAGACTTTGTCCTCTGATGTG 6527

Qy        204 ValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeu------ 221
                                |||   |||               |||   :::   |||
Db       6528 -----------AGTGAGCTGTTAAAAGACAAAACTCATCTCCAGGAAAAGCTGCAGAGT 6575

Qy        222 LeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnValLeu 241
              |||::::::   :::||||||:::                              ||||||:::
Db       6576 TTGGAAAAGGACTCACAGGCACTGTCTTTGACAAAATGTGAGCTGGAAAACCAAATTGCA 6635

Qy        242 MetLeuGlyLys-----------------ThrSerIleSerAspLeuSerGluHisVal 255
                 |||   |||                                     |||            :::
Db       6636 CAACTGAATAAAGAGAAAGAATTGCTTGTCAAGGAATCTGAAAGCCTGCAGGCCAGACTG 6695

Qy        256 PheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuPro 275
                 :::            |||      |||:::            |||         ::::::::
Db       6696 AGTGAATCAGATTATGAAAAGCTGAATGTCTCCAAGGCCTTGGAGGCCGCACTGGTGGAG 6755

Qy        276 GlnLeuGluPheLysLeuLysSerAsnAspAsnGluGlu--------------ArgLeu 290
              :::   ||||||   |||:::   :::         :::|||                |||
Db       6756 AAAGGTGAGTTCGCATTGAGGCTGAGCTCAACACAGGAGGAAGTGCATCAGCTGAGAAGA 6815

Qy        291 GlnValValLysLeuLeuAlaLysMetPheGly-AlaLysAspSerGluLeuAlaSerGl 310
                 :::   ||||||      ::::::               |||   |||         |||||
Db       6816 GGCATCGAGAAACTGAGAGTTCGCATTGAGGCCGATGAAAAGAAGCAGCTGCACATCGCA 6875

Qy        310 nAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIleAr 330
               |                   :::            ||||||            ||||||:::   |||::
Db       6876 GAGAAAC------TGAAAGAACGCGAGCGGGAGAATG------ATTCACTTAAGGATAAA 6923

Qy        330 gLeuGluCysValLysPheAlaSerHisCys--------------------------Le 341
              :|||      :::|||            |||                              ||
Db       6924 GTTGAGAACCTTGAAAGGGAATTGCAGATGTCAGAAGAAAACCAGGAGCTAGTGATTCTT 6983

Qy        341 uMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHisAs 361
              ||||      |||   :::   |||            |||   ||||||   :::         ::
Db       6984 GATG------CCGAGAATTCCAAAGCAGAA-GTAGAGACTCTAAAACACAAATAGAAGA 7036

Qy        361 pProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLysAs 381
              :               :::::::::         :::         :::         :::::|||::
Db       7037 GATGGCCAGAAGCCTGAAGATTTTTGAATTAGACCTTGTCACGTTAAGGTCTGAAAAAGA 7096

Qy        381 pIleLeuLeuValAsn-------------------------------AspHisLeuLe 390
              :   |||                                            |||   |||||
Db       7097 AAATCTGACAAAACAAATACAAGAAAAACAAGGTCAGTTGTCAGAACTAGACAAGTACT 7156

Qy        390 uAsnPheValArgGluArgThrLeuAspLysArgTrp---------ArgValArgLysGl 407
              |:::         :::         :::|||                  :::::::::::||
Db       7157 CTCTTCATTTAAAAGTCTGTTAGAAGAAAAGGAGCAAGCAGAGATACAGATCAAAGAAGA 7216

Qy        407 uAlaMetMetGlyLeuAlaGlnIleTyrLys------LysTyrAlaLeuGlnSerAlaAl 425
              |:::         :::      :::            |||         ::::::   ||
Db       7217 ATCTAAAACTGCAGTGGAGATGCTTCAGAATCAGTTAAAGGAGCTAAATGAGGCAGTAGC 7276

Qy        425 aGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTy 445
              |                  ::::::|||            ::::::   |||
Db       7277 AGCCTTGTGTGGTGACCAAGAAATTATGAAGGCCACAGAACAGAGTCTAGACCCACCAAT 7336

Qy        445 rGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValPr 465
               :::                  |||      :::|||::::::   |||:::
Db       7337 AGAGGAAGAGCATCAGCTGAGAAATAGCATTGAAAAGCTGAGAGCCCGC----------- 7385

Qy        465 oHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAs 485
                   ||||||      |||:::   :::
Db       7386 -------CTAGAAGCTGATGAAAAGAAGCAG--------------------------- 7409

Qy        485 pLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGl 505
```

```
                |||      ::::::   |||   |||                    :::|||||||:::    ::
Db       7410 -CTCTGTGTCTTACAACAACTGAAGGAAAGTGAGCATCATGCAGATTTACTTAAGGGTAG 7468

Qy        505 nValLysAspLeu---------LeuAspLeuIleLysGlnProLysThrAspAlaSerVa 522
              :|||::::::|||          |||::::::  :::        :::       |||:::::
Db       7469 AGTGGAGAACCTTGAAAGAGAGCTAGAGATAGCCAGGACAAACCAAGAGCATGCAGCTCT 7528

Qy        522 lLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAl 542
              ::::|||        ||||||       |||   |||                    |||
Db       7529 TGAGGCAGAGAATTCCAAAGGAGAGGTAGAGACCCTAAAAGCAAAAATAGAAGGGATGAC 7588

Qy        542 aGlnAspPhe------------------------------------MetLysLysPh 549
               |||                                                 :::
Db       7589 CCAAAGTCTGAGAGGTCTGGAATTAGATGTTGTTACTATAAGGTCAGAAAAAGAAAATCT 7648

Qy        549 eThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerPr 569
               |||        |||:::    :::::::::        :::||||||:::::     |||
Db       7649 GACAAATGAATTACAAAAAGAGCAAGAGCGAATATCTGAATTAGAAATAATAAATTCA-- 7706

Qy        569 oThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAs 589
                                     |||      ::::::|||       ::::::
Db       7707 ---------------TCATTTGAAAATATTTTGCAAGAAAAAGAGCAAGAGAAAGTACA 7750

Qy        589 nProLysGlnProThrAsnProPheLeuGluMet------------IleLysPheLeuLe 605
                 |||:::    ::::::       ::::||||||            :::|||    |||
Db       7751 GATGAAAGAAAAAATCAAGCACTGCCATGGAGATGCTTCAAACACAATTAAAAGAGCTCAA 7810

Qy        605 uGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnVa 625
                ||||||::::|||      :::|||      |||       |||::::
Db       7811 TGAGAGAGTGGCAGCCCTGCATAATGACCAAGAAGCCTGTAAGGCC------AAAGAGCA 7864

Qy        625 lAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAl 645
                |||   |||                          :::    |||   :::     ::::::||
Db       7865 GAATCTTAGT-----------------AGTCAAGTAGAGTGTCTTGAACTTGAGAAGGC 7906

Qy        645 a---IleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPh 664
              |    :::        ||||||:::          |||              :::   |||
Db       7907 TCAGTTGCTACAAGGCCTTGATGAGGCCAAA-----------AATAATTATATTGTTTT 7954

Qy        664 eHisSerAla--------------GluThrPheGluSerLeuLeuAlaCysLeuLysMe 679
                |||:::                  :::      |||                |||:::
Db       7955 GCAATCTTCAGTGAAAGGCCTCATTCAAGAAGTAGAAGATGGCAAGCAGAAACTGGAGAA 8014

Qy        679 tAspAspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIl 699
                ||||||:::::::::                ||||||
Db       8015 GAAGGATGAAGAAATCAGTAGACTGAAAAATCAAATT---------------------- 8051

Qy        699 eGluGluAspPheProHisIleArgSerAlaLeuLeuProVal------LeuHisHis-- 716
                :::|||          :::    |||   |||        |||           |||
Db       8052 ----CAAGACCAAGAGCAGCTTGTCTCTAAACTGTCCCAGGTGGAAGGAGAGCACCAACT 8107

Qy        717 ----LysSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAl 735
                   |||   :::             |||                :::
Db       8108 TTGGAAGGAGCAAAACTTAGAACTGAGAAATCTGACGGTGGAATTGGAGCAGAAGATCCA 8167

Qy        735 aIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLe 755
               :::        ||||||                         |||    |||      |||
Db       8168 AGTGCTACAATCCAAAAATGCCTCTTTGCAGGACACATTAGAAGTGCTGCAGAGTTCTTA 8227

Qy        755 uAspProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLe 775
                        |||||||||:::                                :::   ||
Db       8228 C------AAGAATCTAGAGAAT-------------------------GAGCTTGAATT 8254

Qy        775 uLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLy 795
              |         |||:::                                 :::|||:::
Db       8255 GACAAAAATGGACAAAATG--------------------------TCCTTTGTT----- 8282

Qy        795 sAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAs 815
                                               :::|||      |||:::          ::
Db       8283 -----------------------------GAAAAAGTAAACAAAATGACTGCAAAGGA 8311

Qy        815 pGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTr 835
              :   |||:::       |||                    |||             :::
Db       8312 AACTGAGCTGCAGAGGGAAATGCATGAGATGGCACAGAAAACAGCAGAGCTGCAAGAAGA 8371

Qy        835 pLeuLeuGlyMetLysAsn-----------------------------AsnHisSerLy 845
               |||   |||   ||||||                                       ||
Db       8372 ACTCAGTGGAGAGAAAAATAGGCTAGCTGGAGAGTTGCAGTTACTGTTGGAAGAAATAAA 8431

Qy        845 sSerGlyThrSerThrLeuArgLeuLeuThr---------------------------- 855
               ||||                 |||:::    ||||||
Db       8432 GAGCAGCAAAGATCAATTGAAGGAGCTCACACTAGAAAATAGTGAATTGAAGAAGAGCCT 8491

Qy        856 -ThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMe 875
                  :::|||   |||        :::    ::::::||||||:::    :::    :::::
Db       8492 AGATTGCATGCACAAAGAC---CAGGTGGAAAAGGAAGGGAAAGTG---AGAGAGGAAAT 8545

Qy        875 tSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTy 895
               ::::       :::|||                  |||
Db       8546 AGCTGAATATCAGCTACGGCTTCATGAAGCTGAAAAGAAA------------------- 8585

Qy        895 rHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCy 915
               |||:::    :::   |||:::                                ::::::
Db       8586 -CACCAGGCTTTGCTTTTGGAC-------------------------ACAAACAAACA 8617

Qy        915 sTyrGlnValArg---GlnValPheAlaGlnLysLeu---------HisLysGlyLeuSe 931
               |||::::|||      |||    :::    ::::||||||               :::   |||||
Db       8618 GTATGAAGTAGAAATCCAGACATACCGAGAGAAATTGACTTCTAAAGAAGAATGTCTCAG 8677

Qy        931 rArgLeuArgLeuProLeuGluTyrMetAlaIleCys-----------------AlaLe 945
               |       ::::|||    ::::::    :::                       :::||
Db       8678 TTCACAGAAGCTGGAGATAGACCTTTTAAAGTCTAGTAAAGAAGAGCTCAATAATTCATT 8737

Qy        945 uCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAs 965
              |   |||          |||                :::            :::|||
Db       8738 GAAAGCTACTACTCAGATTTTGGAAGAATTGAAGAAAACCAAGATGGACAATCTAAAATA 8797

Qy        965 nIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSe 985
               :::|||                  |||||||:::           ::::::       :::
Db       8798 TGTAAAT-----------CAGTTGAAGAAGGAAAATGAACGTGCCCAGGGGAAAATGAA 8845

Qy        985 rLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTy 1005
               ||||||
Db       8846 GTTGTTG---------------------------------------------------- 8852

Qy       1005 rValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValLe 1025
                    ::::::           :::||||||::::::     ||||||   |||        ||
Db       8853 -------ATCAAATCCTGTAAACAGCTGGAAGAGGAAAAGGAGATACTGCAGAAAGAACT 8905

Qy       1025 uGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArg---------- 1041
              |      |||   |||         |||                  :::
Db       8906 CTCTCAACTTCAAGCTGCACAGGAGAAGCAGAAAACAGGTACTGTTATGGATACCAAGGT 8965

Qy       1042 ----LysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLy 1060
                  ::::::           ||||||:::   |||:::                       ::
Db       8966 CGATGAATTAACAACTGAGATCAAAGAACTGAAAGAAACTCTTGAAGAAAAAACCAAGGA 9025

Qy       1060 sMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSe 1080
              :    ::::|||    |||            |||              ::::::::::::|||
Db       9026 GGCAGATGAATACTTGGATAAGTACTGT-----------TCCTTGCTTATAAGCCATGA 9073

Qy       1080 rThrThrTyrSerLeuGluSerProLysAsp------ProValLeuProAlaArgPhePh 1098
                           ||||||         |||:::                   |||
Db       9074 A---------AAGTTAGAGAAAGCTAAAGAGATGTTAGAGACACAAGTGGCCCATCTGTG 9124

Qy       1098 eThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSe 1118
                ::::|||      |||       |||                |||    |||    :::
Db       9125 TTCACAGCAATCTAAACAAGATTCCCGAGGGTCTCCTTTGCTAGGTCCAGTTGTT----- 9179

Qy       1118 rPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLe 1138
                                  |||       :::    :::                |||   ||
Db       9180 ------------------CCAGGACCATCTCCAATCCCTTCTGTTACTGAAAAGAGGTT 9220

Qy       1138 uSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAl 1158
               |||||||          ::::::        |||        |||            :::
Db       9221 ATCATCTGGCCAAAATAAAGCTTCAGGCAAGAGGCAAAGATCCAGTGGAATATGGGAGAA 9280
```

```
Qy       1158 aSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGl 1178
                                  |||::::::|||
Db       9281 TGGTAGAGGACCAACACCTGCTACCCCA------------------------------ 9308

Qy       1178 uMetAspHisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSe 1198
                                 |||   ::::::                    ||||||||
Db       9309 ------------------GAGAGCTTTTCT--------------------AAAAAAAG 9328

Qy       1198 rAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLy 1218
               |  |||          |||   :::        :::|||
Db       9329 CAAGAAAGCAGTCATGAGTGGTATTCACCCTGCAGAA---------------------- 9365

Qy       1218 sLysThrProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuVa 1238
                  |||       ||||||   |||    :::    |||:::    ::::::   |||
Db       9366 -GACACGGAAGGTACTGAGTTTGAGCCAGAG---GGACTTCCAGAAGTTGTAAAGAAAGG 9421

Qy       1238 lGlnGluGlnLysProLysGly------SerGlnArgSerArgLysArgGlyHisThrAl 1256
                         |||   |||      |||                :::|||        ||
Db       9422 GTTTGCTGACATCCCGACAGGAAAGACTAGCCCATATATCCTGCGAAGAACAACCATGGC 9481

Qy       1256 aSerGluSerAspGluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeu-- 1275
              |:::   :::       :::          ::::::     |||          |||
Db       9482 AACTCGGACCAGCCCCCGCCTGGCTGCACAGAAGTTAGCGCTATCCCCACTGAGTCTCGG 9541

Qy       1276 ----GluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgProPr 1294
                   ||||||      |||   :::   |||          |||   |||    :::
Db       9542 CAAAGAAAATCTTGCAGAGTCCTCCAAACCAACAGCTGGTGGCAGCAGATCACAAAAGGT 9601

Qy       1294 oLys---------------ProLeuGlyGlyGlyThrProLysGluGluProThrMetLy 1309
                |||              |||:::      ||||||        |||||||||    ||
Db       9602 CAAAGTTGCTCAGCGGAGCCCAGTAGATTCAGGCACCATCCTCCGAGAACCCACCACGAA 9661

Qy       1309 s----------------------------------------ThrSerLysLysGlySe 1315
              |                                        :::   ::::::|||
Db       9662 ATCCGTCCCAGTCAATAATCTTCCTGAGAGAAGTCCGACTGACAGCCCCAGAGAGGGCCT 9721

Qy       1315 rLysLysLysSerGly------ProProAlaProGlu 1325
                :::   |||   |||      |||||||||:::
Db       9722 GAGGGTCAAGCGAGGCCGACTTGTCCCAGCCCCAAAG 9758
```

RESULT 6
PCT-US95-16216-2

```
; Sequence 2, Application PC/TUS9516216
;  GENERAL INFORMATION:
;    APPLICANT:  Yen, Timothy J.
;    APPLICANT:  Rattner, Jerome B.
;    TITLE OF INVENTION:  Nucleic Acid Encoding a Transiently
;    TITLE OF INVENTION:  Expressed Kinetochore Protein, and Methods of Use
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dann, Dorfman, Herrell and Skillman
;      STREET:  1601 Market Street   Suite 720
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103-2307
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/16216
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/353,700
;      FILING DATE:  09-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reed, Janet E.
;      REGISTRATION NUMBER:  36,252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-4100
;      TELEFAX:  (215) 563-4044
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10136 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US95-16216-2
```

```
Alignment Scores:
Pred. No.:                6.44e-08     Length:          10136
Score:                    201.50       Matches:         283
Percent Similarity:       35.16%       Conservative:    242
Best Local Similarity:    18.96%       Mismatches:      587
Query Match:              2.80%        Indels:          382
DB:                       5            Gaps:            63
```

US-09-512-581B-2 (1-1391) x PCT-US95-16216-2 (1-10136)

```
Qy         19 ValLysGluIleSerAspLysIle------------------------SerLysGluGlu 30
                :::||||||:::   |||:::                         :::|||   |||
Db       5868 ATGAAAGAATTAGACTCAAAACTCCATTTACAGGAGGTACAACTAATGACCAAAATTGAA 5927

Qy         31 MetValArgArgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGlu 50
                          |||:::   :::|||                       :::   |||
Db       5928 GCATGCATAGAATTGGAAAAAATAGTTGGGGAACTTAAGAAAGAAAACTCAGATTTAAGT 5987

Qy         51 GluGluLysGluLeuTyr------------------------------------------ 56
              |||:::   |||   :::
Db       5988 GAAAAATTGGAATATTTTTCTTGTGATCACCAGGAGTTACTCCAGAGAGTAGAAACTTCT 6047

Qy         57 ------LeuAsnLeuAlaLeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLys 74
                    ||||||      |||   :::   :::|||          |||          :::
Db       6048 GAAGGCCTCAATTCTGATTTAGAAATGCATGCAGAT---------AAATCATCACGTGAA 6098

Qy         75 AspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGlu 94
              |||:::         ||||||
Db       6099 GATATTGGAGATAATGTGGCC--------------------------------------- 6119

Qy         95 AlaProTyrThrSerProAspLysLeuLysAspIlePheMetPheIleThrArgGlnLeu 114
                                 |||      |||:::   |||:::    :::      :::|||
Db       6120 ---------AAGGTGAATGACAGCTGGAAGGAGAGATTTCTTGATGTGGAAAATGAGCTG 6170

Qy        115 LysGlyLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsn 134
                    :::          |||:::         :::         ||||||      :::
Db       6171 AGTAGGATCAGATCGGAGAAAGCTAGCATTGAGCATGAAGCCCTCTACCTGGAGGCTGAC 6230

Qy        135 IleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSerAsn------Glu 152
              :::      |||::::::         :::|||   |||   ::::::   |||      :::
Db       6231 TTAGAGGTAGTTCAAACAGAGAAGCTATGTTTAGAAAAAGACAATGAAAATAAGCAGAAG 6290

Qy        153 IlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAsnGlyHisAsn------ 170
              :::         |||         |||   ||||||:::   :::   |||
Db       6291 GTTATTGTCTGCCTTGAAGAAGAACTC---TCAGTGGTCACAAGTGAGAGAAACCAGCTT 6347

Qy        170 ------------------------------------------------------------ 170

Db       6348 CGTGGAGAATTAGATACTATGTCAAAAAAAACCACGGCACTGGATCAGTTGTCTGAAAAA 6407

Qy        171 ---------------GlnLysValHisMetHisMetValAspLeuMetSerSerIle--- 184
                             |||::::::      |||         :::   :::      |||
Db       6408 ATGAAGGAGAAAACACAAGAGCTTGAGTCTCATCAAAGTGAGTGTCTCCATTGCATTCAG 6467

Qy        185 IleCysGluGlyAspThrValSerGln---GluLeuLeuAspThrValLeuValAsnLeu 203
              :::   |||   :::            :::   ||||||||   |||:::      ::::::
Db       6468 GTGGCAGAGGCAGAGGTGAAGGAAAAGACGGAACTCCTTCAGACTTTGTCCTCTGATGTG 6527

Qy        204 ValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeu------ 221
                                 |||   |||               |||   :::   |||
```

```
Db       6528 ------------AGTGAGCTGTTAAAAGACAAAACTCATCTCCAGGAAAAGCTGCAGAGT 6575

Qy        222 LeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnValLeu 241
              |||:::::   :::||||||:::                      ||||||:::
Db       6576 TTGGAAAAGGACTCACAGGCACTGTCTTTGACAAAATGTGAGCTGGAAAACCAAATTGCA 6635

Qy        242 MetLeuGlyLys-----------------ThrSerIleSerAspLeuSerGluHisVal 255
                 |||   |||                               |||         :::
Db       6636 CAACTGAATAAAGAGAAAGAATTGCTTGTCAAGGAATCTGAAAGCCTGCAGGCCAGACTG 6695

Qy        256 PheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuPro 275
                 :::           |||     |||:::       |||      :::::::::
Db       6696 AGTGAATCAGATTATGAAAAGCTGAATGTCTCCAAGGCCTTGGAGGCCGCACTGGTGGAG 6755

Qy        276 GlnLeuGluPheLysLeuLysSerAsnAspAsnGluGlu--------------ArgLeu 290
              :::   ||||||   |||:::   :::          :::|||            |||
Db       6756 AAAGGTGAGTTCGCATTGAGGCTGAGCTCAACACAGGAGGAAGTGCATCAGCTGAGAAGA 6815

Qy        291 GlnValValLysLeuLeuAlaLysMetPheGly-AlaLysAspSerGluLeuAlaSerGl 310
              :::   ||||||      ::::::          |||   |||         |||||
Db       6816 GGCATCGAGAAACTGAGAGTTCGCATTGAGGCCGATGAAAAGAAGCAGCTGCACATCGCA 6875

Qy        310 nAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIleAr 330
              |                :::      ||||||      ||||||:::   |||::
Db       6876 GAGAAAC------TGAAAGAACGCGAGCGGGAGAATG------ATTCACTTAAGGATAAA 6923

Qy        330 gLeuGluCysValLysPheAlaSerHisCys-------------------------Le 341
              :|||      :::|||            |||                         ||
Db       6924 GTTGAGAACCTTGAAAGGGAATTGCAGATGTCAGAAGAAAACCAGGAGCTAGTGATTCTT 6983

Qy        341 uMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHisAs 361
              ||||      |||   :::   |||      ||||||      :::      ::
Db       6984 GATG------CCGAGAATTCCAAAGCAGAA-GTAGAGACTCTAAAAACACAAATAGAAGA 7036

Qy        361 pProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLysAs 381
              :         ::::::::         :::      :::      ::::::|||::
Db       7037 GATGGCCAGAAGCCTGAAGATTTTGAATTAGACCTTGTCACGTTAAGGTCTGAAAAAGA 7096

Qy        381 pIleLeuLeuValAsn-------------------------------AspHisLeuLe 390
              :   |||                                         |||   |||||
Db       7097 AAATCTGACAAAACAAATACAAGAAAAACAAGGTCAGTTGTCAGAACTAGACAAGTTACT 7156

Qy        390 uAsnPheValArgGluArgThrLeuAspLysArgTrp---------ArgValArgLysGl 407
              |:::      :::            :::|||           :::::::::::||
Db       7157 CTCTTCATTTAAAAGTCTGTTAGAAGAAAAGGAGCAAGCAGAGATACAGATCAAAGAAGA 7216

Qy        407 uAlaMetMetGlyLeuAlaGlnIleTyrLys------LysTyrAlaLeuGlnSerAlaAl 425
              |:::      :::      :::            |||         ::::::   ||
Db       7217 ATCTAAAACTGCAGTGGAGATGCTTCAGAATCAGTTAAAGGAGCTAAATGAGGCAGTAGC 7276

Qy        425 aGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTy 445
              |               ::::::|||         :::   |||
Db       7277 AGCCTTGTGTGGTGACCAAGAAATTATGAAGGCCACAGAACAGAGTCTAGACCCACCAAT 7336

Qy        445 rGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValPr 465
               :::               |||      :::|||::::::   |||:::
Db       7337 AGAGGAAGAGCATCAGCTGAGAAATAGCATTGAAAAGCTGAGAGCCCGC----------- 7385

Qy        465 oHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAs 485
                    ||||||      |||:::   :::
Db       7386 -------CTAGAAGCTGATGAAAAGAAGCAG---------------------------- 7409

Qy        485 pLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGl 505
              |||         ::::::   |||   |||            :::|||||||:::   ::
Db       7410 -CTCTGTGTCTTACAACAACTGAAGGAAAGTGAGCATCATGCAGATTTACTTAAGGGTAG 7468

Qy        505 nValLysAspLeu---------LeuAspLeuIleLysGlnProLysThrAspAlaSerVa 522
              :|||:::::|||            |||::::::   :::         :::   |||:::::
Db       7469 AGTGGAGAACCTTGAAAGAGAGCTAGAGATAGCCAGGACAAACCAAGAGCATGCAGCTCT 7528

Qy        522 lLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAl 542
              ::::|||      ||||||      |||   |||                  |||
Db       7529 TGAGGCAGAGAATTCCAAAGGAGAGGTAGAGACCCTAAAAGCAAAAATAGAAGGGATGAC 7588

Qy        542 aGlnAspPhe--------------------------------------MetLysLysPh 549
               |||                                             :::
Db       7589 CCAAAGTCTGAGAGGTCTGGAATTAGATGTTGTTACTATAAGGTCAGAAAAAGAAAATCT 7648

Qy        549 eThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerPr 569
               |||      |||:::   :::::::::         :::||||||:::::   |||
Db       7649 GACAAATGAATTACAAAAAGAGCAAGAGCGAATATCTGAATTAGAAATAATAAATTCA-- 7706

Qy        569 oThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAs 589
                                |||      ::::::|||      ::::::
Db       7707 ----------------TCATTTGAAAATATTTTGCAAGAAAAAGAGCAAGAGAAAGTACA 7750

Qy        589 nProLysGlnProThrAsnProPheLeuGluMet-----------IleLysPheLeuLe 605
               |||:::   ::::::      :::||||||           :::|||   |||
Db       7751 GATGAAAGAAAAAATCAAGCACTGCCATGGAGATGCTTCAAACACAATTAAAAGAGCTCAA 7810

Qy        605 uGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnVa 625
               ||||||:::|||   :::|||   |||   ||||||:::      |||   |||:::
Db       7811 TGAGAGAGTGGCAGCCCTGCATAATGACCAAGAAGCCTGTAAGGCC------AAAGAGCA 7864

Qy        625 lAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAl 645
              |||   |||                        :::   |||   :::      :::::||
Db       7865 GAATCTTAGT-----------------AGTCAAGTAGAGTGTCTTGAACTTGAGAAGGC 7906

Qy        645 a---IleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPh 664
              |   :::      ||||||:::      |||                  :::   |||
Db       7907 TCAGTTGCTACAAGGCCTTGATGAGGCCAAA-----------AATAATTATATTGTTTT 7954

Qy        664 eHisSerAla--------------GluThrPheGluSerLeuLeuAlaCysLeuLysMe 679
                |||:::               :::      |||               |||:::
Db       7955 GCAATCTTCAGTGAAAGGCCTCATTCAAGAAGTAGAAGATGGCAAGCAGAAACTGGAGAA 8014

Qy        679 tAspAspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIl 699
                ||||||:::::::::               ||||||
Db       8015 GAAGGATGAAGAAATCAGTAGACTGAAAAATCAAATT--------------------- 8051

Qy        699 eGluGluAspPheProHisIleArgSerAlaLeuLeuProVal------LeuHisHis-- 716
                :::|||         :::   |||   |||         |||            |||
Db       8052 ----CAAGACCAAGAGCAGCTTGTCTCTAAACTGTCCCAGGTGGAAGGAGAGCACCAACT 8107

Qy        717 ----LysSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAl 735
                  |||   :::            |||            :::
Db       8108 TTGGAAGGAGCAAAACTTAGAACTGAGAAATCTGACGGTGGAATTGGAGCAGAAGATCCA 8167

Qy        735 aIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLe 755
               :::         ||||||                     |||   |||      |||
Db       8168 AGTGCTACAATCCAAAAATGCCTCTTTGCAGGACACATTAGAAGTGCTGCAGAGTTCTTA 8227

Qy        755 uAspProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLe 775
                    |||||||||:::                               :::   ||
Db       8228 C------AAGAATCTAGAGAAT----------------------------GAGCTTGAATT 8254

Qy        775 uLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLy 795
              |                  |||:::                             :::|||:::
Db       8255 GACAAAAATGGACAAAATG--------------------------TCCTTTGTT----- 8282

Qy        795 sAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAs 815
                                       :::|||      |||:::             ::
Db       8283 -------------------------------GAAAAAGTAAACAAAATGACTGCAAAGGA 8311

Qy        815 pGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTr 835
              :   |||:::         |||         |||                  :::
Db       8312 AACTGAGCTGCAGAGGGAAATGCATGAGATGGCACAGAAAACAGCAGAGCTGCAAGAAGA 8371

Qy        835 pLeuLeuGlyMetLysAsn-----------------------------AsnHisSerLy 845
               |||   |||   ||||||                                      ||
Db       8372 ACTCAGTGGAGAGAAAAATAGGCTAGCTGGAGAGTTGCAGTTACTGTTGGAAGAAATAAA 8431

Qy        845 sSerGlyThrSerThrLeuArgLeuLeuThr--------------------------- 855
               ||||            |||:::   ||||||
Db       8432 GAGCAGCAAAGATCAATTGAAGGAGCTCACACTAGAAAATAGTGAATTGAAGAAGAGCCT 8491
```

```
Qy        856 -ThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMe 875
              :::|||  |||      :::    :::::||||||:::    :::    :::::
Db       8492 AGATTGCATGCACAAAGAC---CAGGTGGAAAAGGAAGGGAAAGTG---AGAGAGGAAAT 8545

Qy        875 tSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTy 895
              ::::     :::|||          |||       |||
Db       8546 AGCTGAATATCAGCTACGGCTTCATGAAGCTGAAAAGAAA------------------- 8585

Qy        895 rHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCy 915
               |||:::   :::   |||:::                              ::::::
Db       8586 -CACCAGGCTTTGCTTTTGGAC-------------------------ACAAACAAACA 8617

Qy        915 sTyrGlnValArg---GlnValPheAlaGlnLysLeu---------HisLysGlyLeuSe 931
               |||:::|||      |||   :::   :::||||||          :::   |||||
Db       8618 GTATGAAGTAGAAATCCAGACATACCGAGAGAAATTGACTTCTAAAGAAGAATGTCTCAG 8677

Qy        931 rArgLeuArgLeuProLeuGluTyrMetAlaIleCys-----------------AlaLe 945
              |     :::|||   ::::::   :::                           :::||
Db       8678 TTCACAGAAGCTGGAGATAGACCTTTTAAAGTCTAGTAAAGAAGAGCTCAATAATTCATT 8737

Qy        945 uCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAs 965
              |   |||          :::   |||                  :::      :::|||
Db       8738 GAAAGCTACTACTCAGATTTTGGAAGAATTGAAGAAAACCAAGATGGACAATCTAAAATA 8797

Qy        965 nIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSe 985
               :::|||              ||||||:::          ::::::      :::
Db       8798 TGTAAAT-----------CAGTTGAAGAAGGAAAATGAACGTGCCCAGGGGAAAATGAA 8845

Qy        985 rLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTy 1005
               ||||||
Db       8846 GTTGTTG----------------------------------------------------- 8852

Qy       1005 rValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValLe 1025
                 ::::::      :::||||||||::::::   ||||||   |||          ||
Db       8853 -------ATCAAATCCTGTAAACAGCTGGAAGAGGAAAAGGAGATACTGCAGAAAGAACT 8905

Qy       1025 uGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArg----------- 1041
              |     |||   |||         |||               :::
Db       8906 CTCTCAACTTCAAGCTGCACAGGAGAAGCAGAAAACAGGTACTGTTATGGATACCAAGGT 8965

Qy       1042 ----LysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLy 1060
                  ::::::         ||||||:::   |||:::                   ::
Db       8966 CGATGAATTAACAACTGAGATCAAAGAACTGAAAGAAACTCTTGAAGAAAAAACCAAGGA 9025

Qy       1060 sMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSe 1080
              :    :::|||   |||         |||            ::::::::::::|||
Db       9026 GGCAGATGAATACTTGGATAAGTACTGT-----------TCCTTGCTTATAAGCCATGA 9073

Qy       1080 rThrThrTyrSerLeuGluSerProLysAsp------ProValLeuProAlaArgPhePh 1098
                        ||||||      |||:::                   |||
Db       9074 A---------AAGTTAGAGAAAGCTAAAGAGATGTTAGAGACACAAGTGGCCCATCTGTG 9124

Qy       1098 eThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSe 1118
               :::|||      |||      |||               |||   |||   :::
Db       9125 TTCACAGCAATCTAAACAAGATTCCCGAGGGTCTCCTTTGCTAGGTCCAGTTGTT----- 9179

Qy       1118 rPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLe 1138
                                  |||      :::   :::                |||  ||
Db       9180 ------------------CCAGGACCATCTCCAATCCCTTCTGTTACTGAAAAGAGGTT 9220

Qy       1138 uSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAl 1158
              |||||||         ::::::      |||      |||      :::
Db       9221 ATCATCTGGCCAAAATAAAGCTTCAGGCAAGAGGCAAAGATCCAGTGGAATATGGGAGAA 9280

Qy       1158 aSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGl 1178
                                |||::::::|||
Db       9281 TGGTAGAGGACCAACACCTGCTACCCCA-------------------------------- 9308

Qy       1178 uMetAspHisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSe 1198
                                  |||   ::::::                      ||||||||
Db       9309 ------------------GAGAGCTTTTCT--------------------AAAAAAAG 9328

Qy       1198 rAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLy 1218
                |   |||         |||   :::         :::|||
Db       9329 CAAGAAAGCAGTCATGAGTGGTATTCACCCTGCAGAA---------------------- 9365

Qy       1218 sLysThrProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuVa 1238
                 |||      ||||||   |||   :::   |||:::   ::::::   |||
Db       9366 -GACACGGAAGGTACTGAGTTTGAGCCAGAG---GGACTTCCAGAAGTTGTAAAGAAAGG 9421

Qy       1238 lGlnGluGlnLysProLysGly------SerGlnArgSerArgLysArgGlyHisThrAl 1256
                        |||   |||      |||         :::|||          ||
Db       9422 GTTTGCTGACATCCCGACAGGAAAGACTAGCCCATATATCCTGCGAAGAACAACCATGGC 9481

Qy       1256 aSerGluSerAspGluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeu-- 1275
              |:::   :::        :::         ::::::      |||            |||
Db       9482 AACTCGGACCAGCCCCCGCCTGGCTGCACAGAAGTTAGCGCTATCCCCACTGAGTCTCGG 9541

Qy       1276 ----GluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgProPr 1294
                  ||||||      |||   :::   |||         |||   |||   :::
Db       9542 CAAAGAAAATCTTGCAGAGTCCTCCAAACCAACAGCTGGTGGCAGCAGATCACAAAAGGT 9601

Qy       1294 oLys---------------ProLeuGlyGlyGlyThrProLysGluGluProThrMetLy 1309
                |||                |||:::      ||||||         |||||||||  ||
Db       9602 CAAAGTTGCTCAGCGGAGCCCAGTAGATTCAGGCACCATCCTCCGAGAACCCACCACGAA 9661

Qy       1309 s-----------------------------------------ThrSerLysLysGlySe 1315
              |                                         :::   ::::::|||
Db       9662 ATCCGTCCCAGTCAATAATCTTCCTGAGAGAAGTCCGACTGACAGCCCCAGAGAGGGCCT 9721

Qy       1315 rLysLysLysSerGly------ProProAlaProGlu 1325
               :::   |||   |||         |||||||||:::
Db       9722 GAGGGTCAAGCGAGGCCGACTTGTCCCAGCCCCAAAG 9758
```

RESULT 7
US-08-328-254-5
; Sequence 5, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8789 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

```
;     FEATURE:
;        NAME/KEY:  CDS
;        LOCATION:  544..7990
US-08-328-254-5

Alignment Scores:
Pred. No.:              1.36e-07        Length:         8789
Score:                  197.00          Matches:        280
Percent Similarity:     35.77%          Conservative:   248
Best Local Similarity:  18.97%          Mismatches:     604
Query Match:            2.74%           Indels:         345
DB:                     1               Gaps:           62

US-09-512-581B-2 (1-1391) x US-08-328-254-5 (1-8789)

Qy         19 ValLysGluIleSerAspLysIle-----------------------SerLysGluGlu 30
              :::|||||:::         |||:::                       :::|||   |||
Db       4057 ATGAAAGAATTAGACTCAAAACTCCATTTACAGGAGGTACAACTAATGACCAAAATTGAA 4116

Qy         31 MetValArgArgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGlu 50
                                |||:::   :::|||                 :::   |||
Db       4117 GCATGCATAGAATTGGAAAAAATAGTTGGGGAACTTAAGAAAGAAAACTCAGATTTAAGT 4176

Qy         51 GluGluLysGluLeuTyr------------------------------------------ 56
              |||:::   |||   :::
Db       4177 GAAAAATTGGAATATTTTTCTTGTGATCACCAGGAGTTACTCCAGAGAGTAGAAACTTCT 4236

Qy         57 ------LeuAsnLeuAlaLeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLys 74
                    ||||||   |||   :::   :::|||            |||         :::
Db       4237 GAAGGCCTCAATTCTGATTTAGAAATGCATGCAGAT---------AAATCATCACGTGAA 4287

Qy         75 AspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGlu 94
              |||:::            ||||||
Db       4288 GATATTGGAGATAATGTGGCC--------------------------------------- 4308

Qy         95 AlaProTyrThrSerProAspLysLeuLysAspIlePheMetPheIleThrArgGlnLeu 114
                                |||      |||:::   |||:::   :::      :::|||
Db       4309 ---------AAGGTGAATGACAGCTGGAAGGAGAGATTTCTTGATGTGGAAAATGAGCTG 4359

Qy        115 LysGlyLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsn 134
                 :::         |||:::            :::      ||||||         :::
Db       4360 AGTAGGATCAGATCGGAGAAAGCTAGCATTGAGCATGAAGCCCTCTACCTGGAGGCTGAC 4419

Qy        135 IleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSerAsn------Glu 152
              :::      |||::::::         :::|||   |||      ::::::         :::
Db       4420 TTAGAGGTAGTTCAAACAGAGAAGCTATGTTTAGAAAAAGACAATGAAAATAAGCAGAAG 4479

Qy        153 IlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAsnGlyHisAsn------ 170
              :::         |||         |||   ||||||:::   :::         |||
Db       4480 GTTATTGTCTGCCTTGAAGAAGAACTC---TCAGTGGTCACAAGTGAGAGAAACCAGCTT 4536

Qy        170 ------------------------------------------------------------ 170

Db       4537 CGTGGAGAATTAGATACTATGTCAAAAAAAACCACGGCACTGGATCAGTTGTCTGAAAAA 4596

Qy        171 --------------GlnLysValHisMetHisMetValAspLeuMetSerSerIle--- 184
                            |||::::::         |||      :::   :::      |||
Db       4597 ATGAAGGAGAAAACACAAGAGCTTGAGTCTCATCAAAGTGAGTGTCTCCATTGCATTCAG 4656

Qy        185 IleCysGluGlyAspThrValSerGln---GluLeuLeuAspThrValLeuValAsnLeu 203
              :::   |||      :::         :::   |||||||||   |||:::      ::::::
Db       4657 GTGGCAGAGGCAGAGGTGAAGGAAAAGACGGAACTCCTTCAGACTTTGTCCTCTGATGTG 4716

Qy        204 ValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeu------ 221
                                |||   |||               |||   :::   |||
Db       4717 ------------AGTGAGCTGTTAAAAGACAAAACTCATCTCCAGGAAAAGCTGCAGAGT 4764

Qy        222 LeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnValLeu 241
              |||::::::   :::||||||:::                     ||||||:::
Db       4765 TTGGAAAAGGACTCACAGGCACTGTCTTTGACAAAATGTGAGCTGGAAAACCAAATTGCA 4824

Qy        242 MetLeuGlyLys-----------------ThrSerIleSerAspLeuSerGluHisVal 255
                 |||   |||                                     |||         :::
Db       4825 CAACTGAATAAAGAGAAAGAATTGCTTGTCAAGGAATCTGAAAGCCTGCAGGCCAGACTG 4884

Qy        256 PheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuPro 275
                 :::         |||      |||:::         |||         :::::::::
Db       4885 AGTGAATCAGATTATGAAAAGCTGAATGTCTCCAAGGCCTTGGAGGCCGCACTGGTGGAG 4944

Qy        276 GlnLeuGluPheLysLeuLysSerAsnAspAsnGluGlu--------------ArgLeu 290
              :::   ||||||   |||:::   :::      :::|||                  |||
Db       4945 AAAGGTGAGTTCGCATTGAGGCTGAGCTCAACACAGGAGGAAGTGCATCAGCTGAGAAGA 5004

Qy        291 GlnValValLysLeuLeuAlaLysMetPheGly-AlaLysAspSerGluLeuAlaSerGl 310
                 :::   ||||||      ::::::            |||   |||         |||||
Db       5005 GGCATCGAGAAACTGAGAGTTCGCATTGAGGCCGATGAAAAGAAGCAGCTGCACATCGCA 5064

Qy        310 nAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIleAr 330
              |                 :::            ||||||         ||||||:::   |||:::
Db       5065 GAGAAAC------TGAAAGAACGCGAGCGGGAGAATG------ATTCACTTAAGGATAAA 5112

Qy        330 gLeuGluCysValLysPheAlaSerHisCys-------------------------Le 341
               :|||      :::|||            |||                        ||
Db       5113 GTTGAGAACCTTGAAAGGGAATTGCAGATGTCAGAAGAAAACCAGGAGCTAGTGATTCTT 5172

Qy        341 uMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHisAs 361
              ||||      |||   :::   |||         |||   ||||||   :::      ::
Db       5173 GATG------CCGAGAATTCCAAAGCAGAA-GTAGAGACTCTAAAAACACAAATAGAAGA 5225

Qy        361 pProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLysAs 381
              :            :::::::::         :::         :::         ::::::|||:::
Db       5226 GATGGCCAGAAGCCTGAAAGTTTTGAATTAGACCTTGTCACGTTAAGGTCTGAAAAAGA 5285

Qy        381 pIleLeuLeuValAsn--------------------------------AspHisLeuLe 390
              :   |||                                         |||   |||||
Db       5286 AAATCTGACAAAACAAATACAAGAAAAACAAGGTCAGTTGTCAGAACTAGACAAGTTACT 5345

Qy        390 uAsnPheValArgGluArgThrLeuAspLysArgTrp---------ArgValArgLysGl 407
              |:::      :::            :::|||            ::::::::::::|||
Db       5346 CTCTTCATTTAAAAGTCTGTTAGAAGAAAAGGAGCAAGCAGAGATACAGATCAAAGAAGA 5405

Qy        407 uAlaMetMetGlyLeuAlaGlnIleTyrLys------LysTyrAlaLeuGlnSerAlaAl 425
              |:::         :::   :::               |||         ::::::   ||
Db       5406 ATCTAAAACTGCAGTGGAGATGCTTCAGAATCAGTTAAAGGAGCTAAATGAGGCAGTAGC 5465

Qy        425 aGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTy 445
              |                  ::::::|||               ::::::   |||
Db       5466 AGCCTTGTGTGGTGACCAAGAAATTATGAAGGCCACAGAACAGAGTCTAGACCCACCAAT 5525

Qy        445 rGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValPr 465
                :::               |||      :::|||:::::      |||:::
Db       5526 AGAGGAAGAGCATCAGCTGAGAAATAGCATTGAAAAGCTGAGAGCCCGC----------- 5574

Qy        465 oHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAs 485
                    ||||||      |||:::   :::
Db       5575 -------CTAGAAGCTGATGAAAAGAAGCAG----------------------------- 5598

Qy        485 pLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGl 505
               |||         ::::::   |||   |||                  :::|||||||:::   ::
Db       5599 -CTCTGTGTCTTACAACAACTGAAGGAAAGTGAGCATCATGCAGATTTACTTAAGGGTAG 5657

Qy        505 nValLysAspLeu---------LeuAspLeuIleLysGlnProLysThrAspAlaSerVa 522
              :|||:::::::|||          |||::::::   :::         :::      |||::::::
Db       5658 AGTGGAGAACCTTGAAAGAGAGCTAGAGATAGCCAGGACAAACCAAGAGCATGCAGCTCT 5717

Qy        522 lLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAl 542
              ::::|||            ||||||      |||   |||                  |||
Db       5718 TGAGGCAGAGAATTCCAAAGGAGAGGTAGAGACCCTAAAAGCAAAAATAGAAGGGATGAC 5777

Qy        542 aGlnAspPhe----------------------------------------MetLysLysPh 549
               |||                                                  :::
Db       5778 CCAAAGTCTGAGAGGTCTGGAATTAGATGTTGTTACTATAAGGTCAGAAAAAGAAAATCT 5837

Qy        549 eThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerPr 569
               |||      |||:::   :::::::::            :::||||||::::::   |||
Db       5838 GACAAATGAATTACAAAAAGAGCAAGAGCGAATATCTGAATTAGAAATAATAAATTCA-- 5895
```

```
Qy        569 oThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAs 589
                                               |||    ::::::|||      ::::::
Db       5896 ---------------TCATTTGAAAATATTTGCAAGAAAAAGAGCAAGAGAAAGTACA 5939

Qy        589 nProLysGlnProThrAsnProPheLeuGluMet-----------IleLysPheLeuLe 605
                 |||:::   ::::::      :::||||||              :::|||   |||
Db       5940 GATGAAAGAAAAATCAAGCACTGCCATGGAGATGCTTCAAACACAATTAAAAGAGCTCAA 5999

Qy        605 uGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnVa 625
                ||||||:::|||   :::|||   |||   |||:::      |||      |||:::
Db       6000 TGAGAGAGTGGCAGCCCTGCATAATGACCAAGAAGCCTGTAAGGCC------AAAGAGCA 6053

Qy        625 lAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAl 645
                |||   |||                  :::   |||   :::      ::::::||
Db       6054 GAATCTTAGT-----------------AGTCAAGTAGAGTGTCTTGAACTTGAGAAGGC 6095

Qy        645 a---IleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPh 664
              |    :::      ||||||:::      |||            :::   |||
Db       6096 TCAGTTGCTACAAGGCCTTGATGAGGCCAAA-----------AATAATTATATTGTTTT 6143

Qy        664 eHisSerAla--------------GluThrPheGluSerLeuLeuAlaCysLeuLysMe 679
                |||:::                 :::      |||            |||:::
Db       6144 GCAATCTTCAGTGAATGGCCTCATTCAAGAAGTAGAAGATGCAAGCAGAAACTGGAGAA 6203

Qy        679 tAspAspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIl 699
                ||||||:::::::::            ||||||
Db       6204 GAAGGATGAAGAAATCAGTAGACTGAAAAATCAAATT---------------------- 6240

Qy        699 eGluGluAspPheProHisIleArgSerAlaLeuLeuProVal------LeuHisHis-- 716
                 :::|||      :::   |||   |||   |||      |||
Db       6241 ----CAAGACCAAGAGCAGCTTGTCTCTAAACTGTCCCAGGTGGAAGGAGAGCACCAACT 6296

Qy        717 ----LysSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAl 735
                  |||   :::         |||            :::
Db       6297 TTGGAAGGAGCAAAACTTAGAACTGAGAAATCTGACAGTGGAATTGGAGCAGAAGATCCA 6356

Qy        735 aIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLe 755
                :::      ||||||                     |||   |||      |||
Db       6357 AGTGCTACAATCCAAAAATGCCTCTTTGCAGGACACATTAGAAGTGCTGCAGAGTTCTTA 6416

Qy        755 uAspProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLe 775
                          |||||||||:::                          :::   ||
Db       6417 C------AAGAATCTAGAGAAT--------------------------GAGCTTGAATT 6443

Qy        775 uLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLy 795
              |   |||:::                                      :::|||:::
Db       6444 GACAAAAATGGACAAAATG-------------------------TCCTTTGTT----- 6471

Qy        795 sAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAs 815
                                            ::::|||   |||:::            ::
Db       6472 ------------------------------GAAAAAGTAAACAAAATGACTGCAAAGGA 6500

Qy        815 pGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTr 835
              :   |||:::      |||               |||            :::
Db       6501 AACTGAGCTGCAGAGGGAAATGCATGAGATGGCACAGAAAACAGCAGAGCTGCAAGAAGA 6560

Qy        835 pLeuLeuGlyMetLysAsn----------------------------AsnHisSerLy 845
                |||   |||   ||||||                                     ||
Db       6561 ACTCAGTGGAGAGAAAAATAGGCTAGCTGGAGAGTTGCAGTTACTGTTGGAAGAAATAAA 6620

Qy        845 sSerGlyThrSerThrLeuArgLeuLeuThr---------------------------- 855
              ||||          |||:::   ||||||
Db       6621 GAGCAGCAAAGATCAATTGAAGGAGCTCACACTAGAAAATAGTGAATTGAAGAAGAGCCT 6680

Qy        856 -ThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMe 875
                 :::|||   |||      :::   ::::::||||||:::   :::   :::::
Db       6681 AGATTGCATGCACAAAGAC---CAGGTGGAAAAGGAAGGGAAAGTG---AGAGAGGAAAT 6734

Qy        875 tSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTy 895
              ::::      :::|||            |||      |||
Db       6735 AGCTGAATATCAGCTACGGCTTCATGAAGCTGAAAAGAAA------------------- 6774

Qy        895 rHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCy 915
               |||:::   :::   |||:::                              ::::::
Db       6775 -CACCAGGCTTTGCTTTTGGAC-------------------------ACAAACAAACA 6806

Qy        915 sTyrGlnValArg---GlnValPheAlaGlnLysLeu---------HisLysGlyLeuSe 931
               |||:::|||      |||   :::   :::||||||            :::   |||||
Db       6807 GTATGAAGTAGAAATCCAGACATACCGAGAGAAATTGACTTCTAAAGAAGAATGTCTCAG 6866

Qy        931 rArgLeuArgLeuProLeuGluTyrMetAlaIleCys-----------------AlaLe 945
              |      :::|||   ::::::   :::                            :::||
Db       6867 TTCACAGAAGCTGGAGATAGACCTTTTAAAGTCTAGTAAAGAAGAGCTCAATAATTCATT 6926

Qy        945 uCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAs 965
              |   |||            :::   |||                  :::      :::|||
Db       6927 GAAAGCTACTACTCAGATTTTGGAAGAATTGAAGAAAACCAAGATGGACAATCTAAAATA 6986

Qy        965 nIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSe 985
              :::|||               ||||||:::         ::::::      :::
Db       6987 TGTAAAT-----------CAGTTGAAGAAGGAAAATGAACGTGCCCAGGGGAAAATGAA 7034

Qy        985 rLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTy 1005
              ||||||
Db       7035 GTTGTTG---------------------------------------------------- 7041

Qy       1005 rValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValLe 1025
                     ::::::      :::|||||||::::::   ||||||   |||         ||
Db       7042 -------ATCAAATCCTGTAAACAGCTGGAAGAGGAAAAGGAGATACTGCAGAAAGAACT 7094

Qy       1025 uGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArg---------- 1041
              |      |||   |||         |||                  :::
Db       7095 CTCTCAACTTCAAGCTGCACAGGAGAAGCAGAAAACAGGTACTGTTATGGATACCAAGGT 7154

Qy       1042 ----LysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLy 1060
                  ::::::               ||||||:::   |||:::                ::
Db       7155 CGATGAATTAACAACTGAGATCAAAGAACTGAAAGAAACTCTTGAAGAAAAAACCAAGGA 7214

Qy       1060 sMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSe 1080
              :   :::|||   |||               :::::::::::|||
Db       7215 GGCAGATGAATACTTGGATAAGTACTGT------------TCCTTGCTTATAAGCCATGA 7262

Qy       1080 rThrThrTyrSerLeuGluSerProLysAsp------ProValLeuProAlaArgPhePh 1098
                       ||||||   |||:::                              |||
Db       7263 A---------AAGTTAGAGAAAGCTAAAGAGATGTTAGAGACACAAGTGGCCCATCTGTG 7313

Qy       1098 eThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSe 1118
                :::|||   |||   |||                  |||   |||   :::
Db       7314 TTCACAGCAATCTAAACAAGATTCCCGAGGGTCTCCTTTGCTAGGTCCAGTTGTT----- 7368

Qy       1118 rPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLe 1138
                                  |||      :::   :::                  |||   ||
Db       7369 -----------------CCAGGACCATCTCCAATCCCTTCTGTTACTGAAAAGAGGTT 7409

Qy       1138 uSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAl 1158
              ||||||                  |||            |||         :::
Db       7410 ATCATCTGGCCAAAATAAAGCTTCAGGCAAGAGGCAAAGATCCAGTGGAATATGGGAGAA 7469

Qy       1158 aSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGl 1178
                                |||:::::::|||               :::         :::
Db       7470 TGGTGGAGGACCAACACCTGCTACCCCAGAGAGCTTTTCTAAAAAAAGCAAGAAAGCAGT 7529

Qy       1178 uMetAsp---------HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGl 1195
                |||                  :::|||:::   :::   |||         |||
Db       7530 CATGAGTGGTATTCACCCTGCAGAAGACACGGAAGGTACTGAGTTTGAGCCAGAGGGACT 7589

Qy       1195 yLysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysProAr 1215
                 :::      |||:::         :::|||:::         :::
Db       7590 TCCAGAAGTTGTAAAGAAAGGGTTTGCTGACATCCCGACAGGAAAGACTAGCCCATATAT 7649

Qy       1215 gGlyArgLysLysThrProValThrGluGlnGlu--------------GluLysLeuGl 1230
               |||:::   |||   |||                               :::||||||
Db       7650 CCTGCGAAGAACAACCATGGCAACTCGGACCAGCCCCCGCCTGGCTGCACAGAAGTTAGC 7709

Qy       1230 yMetAspAspLeuThr-----------LysLeuValGlnGluGlnLysPro---LysGl 1245
```

```
                :::          |||:::                      |||    :::        |||||        ||
Db       7710 GCTATCCCCACTGAGTCTCGGCAAAGAAAATCTTGCAGAGTCCTCCAAACCAACAGCTGG 7769

Qy       1245 ySerGlnArgSerArgLys---ArgGlyHisThrAlaSerGluSerAspGluGlnGlnTr 1264
                |       ||||||:::|||    :::                |||         |||
Db       7770 TGGCAGCAGATCACAAAAGGTCAAAGTTGCTCAGCGGAGCCCAGTAGATTCAGGCACCAT 7829

Qy       1264 pProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsn----- 1282
                       |||           |||           :::    |||        |||::::::
Db       7830 CCTCCGAGAACCCACCACGAAATCCGTCCCAGTCAATAATCTTCCTGAGAGAAGTCCGAC 7889

Qy       1283 -SerProProLysLysGly---------LysArgGlyArg---ProProLysProLeuGl 1298
                       |||:::::|||             |||||||||||           |||    |||
Db       7890 TGACAGCCCCAGAGAGGGCCTGAGGGTCAAGCGAGGCCGACTTGTCCCCAGCCCCAAAGC 7949

Qy       1298 yGly-GlyThrProLysGluGluProThrMetLysThrSerLys 1312
                 |||     :::|||:::               |||::::::::::|||:::
Db       7950 TGGACTGGAGTCCAAGGGCAGTGAGAACTGTAAGGTCCAGTGAA 7993
```

RESULT 8
US-09-091-501B-9
; Sequence 9, Application US/09091501B
; Patent No. 6518413
; GENERAL INFORMATION:
; APPLICANT: Tinsley, Jonathon M
; APPLICANT: Davies, Kay E
; TITLE OF INVENTION: Utrophin gene expression
; FILE REFERENCE: 620-42
; CURRENT APPLICATION NUMBER: US/09/091,501B
; CURRENT FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: PCT/GB96/03156
; PRIOR FILING DATE: 1996-12-19
; PRIOR APPLICATION NUMBER: GB 9525962.8
; PRIOR FILING DATE: 1995-12-19
; PRIOR APPLICATION NUMBER: GB 9615797.9
; PRIOR FILING DATE: 1996-07-26
; PRIOR APPLICATION NUMBER: GB 9622174.2
; PRIOR FILING DATE: 1996-10-24
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
;   LENGTH: 10320
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (11)..(10312)
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Full length
;   OTHER INFORMATION: utrophin construct
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (724)..(758)
;   OTHER INFORMATION: Precise residue is left open
US-09-091-501B-9

Alignment Scores:

Pred. No.:	6.87e-07	Length:	10320
Score:	191.00	Matches:	307
Percent Similarity:	33.86%	Conservative:	232
Best Local Similarity:	19.28%	Mismatches:	553
Query Match:	2.66%	Indels:	502
DB:	4	Gaps:	76

US-09-512-581B-2 (1-1391) x US-09-091-501B-9 (1-10320)

```
Qy          5 LysThrArgThrAsnAspGlyLysIleThrTyrProProGlyValLysGluIleSerAsp 24
                :::||||||        |||    :::                      ||||||||||:::
Db       3296 CAGACAAGACTAGGTGACTACCAAACTCAACTGGAGAAACTTAGCAAGGAGATCGCTACT 3355

Qy         25 LysIleSerLys-----------GluGluMetValArgArgLeuLysMetValValLys 40
                :::    |||:::                :::|||           ||||||       :::
Db       3356 CAAAAAAGTAGGTTGTCTGAAAGTCAAGAAAAAGCTGCGAACCTGAAGAAAGACTTGGCA 3415

Qy         41 ThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAla 60
                           :::              :::|||||||||
Db       3416 GAGATGCAGGAATGGATGACCCAGGCCGAGGAAGAA------------------------ 3451

Qy         61 LeuHisLeuAlaSerAspPhePheLeuLysHisPro------------------------ 72
                :::|||          ||||||           |||    |||
Db       3452 ---TATTTGGAGCGGGATTTTGAGTACAAGTCACCAGAAGAGCTTGAGAGTGCTGTGGAA 3508

Qy         73 -----------------------------GlyLysAspValArgLeuLeuValAlaCys 82
                                              |||:::||||||:::      :::
Db       3509 GAGATGAAGAGGGCAAAAGAGGATGTGTTGCAGAAGGAGGTGAGAGTGAAGATT------ 3562

Qy         83 CysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLys 102
                   |||    |||         ::::::   |||    :::      |||                :::
Db       3563 ---CTCAAGGACAACATCAAGTTATTAGCTGCCAAGGTGCCCTCTGGTGGCCAGGAGTTG 3619

Qy        103 LeuLysAspIle------------------PheMetPheIleThrArgGlnLeuLysGly 116
                      ::::::                      :::        :::       :::::::|||
Db       3620 ACGTCTGAGCTGAATGTTGTGCTGGAGAATTACCAACTTCTTTGTAATAGAATTCGAGGA 3679

Qy        117 -----------LeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeu 132
                                ||||||:::        |||                ||||||
Db       3680 AAGTGCCACACGCTAGAGGAGGTCTGGTCTTGTTGGATTGAACTGCTTCACTATTTGGAT 3739

Qy        133 GluAsnIleAlaTrp--------------------ValLysSerTyrAsnIleCysPhe 145
                             |||                         :::||||||       :::
Db       3740 CTTGAAACTACCTGGTTAAACACTTTGGAAGAGCGGATGAAGAGCACAGAGGTCCTGCCT 3799

Qy        146 GluLeuGluAspSer---AsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerVal 164
                |||        |||:::    ||||||              |||
Db       3800 GAGAAGACGGATGCTGTCAACGAAGCCCTGGAGTCTCTG--------------------- 3838

Qy        165 IleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSerSerIle 184
                                  :::    |||         |||         |||        :::    |||
Db       3839 -------------------GAATCTGTTCTGCGCCACCCGGCAGATAATCGCACCCAGATT 3880

Qy        185 IleCysGluGlyAspThr---ValSerGlnGluLeuLeuAspThrValLeuValAsnLeu 203
                            |||    |||     :::        :::||||||     ::::::
Db       3881 CGAGAGCTTGGCCAGACTCTGATTGATGGGGGGATCCTGGATGATATAATCAGTGAGAAA 3940

Qy        204 ValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLys 223
                :::    |||        :::        :::                ||||||::::::          |||
Db       3941 CTGGAGGCTTTCAACAGCCGATATGAAGATCTAAGTCACCTGGCAGAGAGC------AAG 3994

Qy        224 ArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnValLeuMetLeu 243
                :::    :::              :::        :::                :::|||:::|||     :::
Db       3995 CAGATTTCTTTGGAAAAGCAACTCCAGGTGCTGCGGGAAACTGACCAGATGCTTCAAGTC 4054

Qy        244 GlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeuGluLeuTyr 263
                   :::     |||:::     :::|||    :::     :::                :::     :::
Db       4055 TTGCAAGAGAGCTTGGGGGAGCTGGACAAACAGCTCACCACATACCTGACTGAC------ 4108

Qy        264 AsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSer 283
                   ||||||:::                         :::||||||            |||:::::::::
Db       4109 AGGATAGATGCTTTC-----------CAAGTTCCACAGGAAGCTCAGAAAATCCAAGCA 4156

Qy        284 AsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPheGlyAlaLys 303
                      :::              |||    :::     :::|||              |||             :::
Db       4157 GAGATCTCAGCCCATGAGCTAACCCTAGAGGAGTTGAGAAGAAATATG---------CGT 4207

Qy        304 AspSerGluLeuAlaSer-GlnAsnLysProLeu-----------------TrpGlnCy 317
                             |||    |||  |||             |||                       |||    ||
Db       4208 TCTCAGCCCCTGACCTCCCCAGAGAGTAGGACTGCCAGAGGAGGAAGTCAGATGGATGTG 4267

Qy        317 sTyrLeuGlyArgPheAsnAspIleHisValProIleArg------------------- 330
                |||||    |||              ::::::               :::
Db       4268 CTACAGAGGAAA-CTCCGAGAGGTGTCCACAAAGTTCCAGCTTTTCCAGAAGCCAGCTAA 4326

Qy        331 ----------------LeuGluCys------------------ValLysPheAlaSerHi 339
                                |||:::|||                       ||||||          ||
Db       4327 CTTCGAGCAGCGCATGCTGGACTGCAAGCGTGTGCTGGATGGCGTGAAAGCAGAACTTCA 4386
```

```
Qy        339 sCysLeu-----------------------------MetAsnHisProAsp-------- 346
              |   |||                                     |||    |||
Db       4387 CGTTCTGGATGTGAAGGACGTAGACCCTGACGTCATACAGACGCACCTGGACAAGTGTAT 4446

Qy        347 ----LeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAl 365
                  |||   |||   |||:::|||    :::|||:::                 :::
Db       4447 GAAACTGTATAAAACTTTGAGTGAA---GTCAAACTTGAAGTGGAAACTGTGATTAAAAC 4503

Qy        365 aIleArgHisAspVal------------------------------------------Il 371
                  ||||||   |||                                          ::
Db       4504 AGGAAGACATATTGTCCAGAAACAGCAAACGGACAACCCAAAAGGGATGGATGAGCAGCT 4563

Qy        371 eValSerIle-----------------------------ValThrAlaAlaLysLysAs 381
              :   |||:::                              ||||||    |||:::||
Db       4564 GACTTCCCTGAAGGTTCTTTACAATGACCTGGGCGCACAGGTGACAGAAGGAAAACAGGA 4623

Qy        381 pIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLysAr 401
              |:::         :::                  :::|||
Db       4624 TCTGGAAAGAGCATCACAGTTGGCCCGGAAAATGAAGAAAGAGGCTGCTTCTCTCTCTGA 4683

Qy        401 gTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLe 421
              |||       |||                  ::::::  :::|||
Db       4684 ATGGCTTTCTGCTACTGAA---------------ACTGAATTGGTACAGAAGTCCACTTC 4728

Qy        421 uGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLe 441
               :::        |||   |||        :::|||:::|||  |||:::    |||
Db       4729 AGAAGGTCTGCTTGGT---GACTTGGATACAGAAATTTCCTGGGCTAAAAATGTTCTGAA 4785

Qy        441 uHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGl 461
                 :::     :::   :::       :::   :::                      ||
Db       4786 GGATCTGGAAAAGAGAAAAGCTGATTTAAATACCATCACAGAGAGTAGTGCTGCCCTGCA 4845

Qy        461 nTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTy 481
              |   ::::::        :::             |||       |||:::
Db       4846 AAACTTGATTGAGGGCAGTGAGCCTATTTTAGAAGAGAGGCTCTGCGTC----------- 4894

Qy        481 rAlaThrLeuAspLeuAsnAla-----------ValLysAlaLeuAsnGluMetTrpLy 497
                          |||||||||             |||:::       |||  .  |||
Db       4895 ------------CTTAACGCTGGGTGGAGCCGAGTTCGTACCTGGACTGAAGATTGG-- 4939

Qy        497 sCysGlnAsnLeuLeuArgHis--------------------------------------- 504
                |||       |||:::   |||
Db       4940 -TGCAATACCTTGATGAACCATCAGAACCAGCTAGAAATATTTGATGGGAACGTGGCTCA 4998

Qy        505 ------------------GlnValLysAspLeuLeuAspLeuIleLysGlnProLysTh 518
                                |||    :::   |||||||||   |||::::::      ||
Db       4999 CATAAGTACCTGGCTTTATCAAGCTGAAGCTCTATTGGATGAAATTGAAAAGAAACCAAC 5058

Qy        518 rAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAs 538
              |
Db       5059 A---------------------------------------------------------- 5059

Qy        538 pProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAsp----- 556
                    |||   ::::::   :::|||:::                 |||:::|||
Db       5060 ----AGTAAACAGGAAGAAATTGTGAAGCGTTTAGTATCTGAGCTGGATGATGCCAACCT 5115

Qy        557 -------GluLysIleArgLysGlnLeuGluValLeuValSerProThrCysSerCysLy 574
                     |||   :::|||   |||      :::|||::::::           |||
Db       5116 CCAGGTTGAAAATGTCCGCGATCAAGCCCTTATTTGATGAATGCCCGTGGAAGCTCAAG 5175

Qy        574 sGlnAlaGluGlyCysValArgGluIleThrLys-----------LysLeuGlyAsnPr 590
                                |||||||:::    :::           :::|||
Db       5176 C-----------------AGGGAGCTTGTAGAACCAAAGTTAGCTGAGCTGAATAGGAA 5217

Qy        590 oLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeu------GluArgIl 608
                :::         :::      ::::::      |||      |||:::       |||   ::
Db       5218 CTTTGAAAAGGTGTCTCAACATATCAAAAGTGCCAAATTGCTAATTGCTCAGGAACCATT 5277

Qy        608 eAlaProValHisIleAspThrGluSer------------------------------- 617
              :              :::   ||||||:::
Db       5278 ATACCAATGTTTGGTCACCACTGAAACATTTGAAACTGGTGTGCCTTTCTCTGACTTGGA 5337

Qy        618 ----------------IleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyTh 632
                                |||       ::::::|||    |||    |||    ::::::     ::
Db       5338 AAAATTAGAAAATGACATAGAAAATATGTTAAAATTTGTGGAAAAACACTTGGAATCCAG 5397

Qy        632 rAlaAspAspGlu---------AspGluGlyValProThrAspGlnAlaIleArgAlaGl 649
              :   :::||||||         :::|||            ::::::   ::::::   ||
Db       5398 TGATGAAGATGAAAAGATGGATGAGGAGAGTGCCCAGATTGAGGAAGTTCTACAAAGAGG 5457

Qy        649 yLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAlaGluTh 669
              |   |||:::|||                     |||:::      :::        |||
Db       5458 AGAAGAAATGTTACAT--------------CAACCTATGGAAGATAATAAAAAAGAAAA 5502

Qy        669 rPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGluAlaAlaLe 689
                          |||    |||                  |||:::                 ::
Db       5503 GATCCGTTTGCAATTATTACTTTTGCATACTAGATACAACAAAATTAAGGCAATCCCTAT 5562

Qy        689 uGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHisIleArgSerAl 709
              :|||                  |||:::    :::             |||||||||::
Db       5563 TCAA--------------CAGAGGAAAATGGGTCAACTTGCTTCTGGAATTAGATCATC 5607

Qy        709 aLeuLeuProValLeuHisHisLysSerLysLysGlyProProArgGlnAlaLysTyrAl 729
              :|||||||||                                         |||
Db       5608 ACTTCTTCCT---------------------------------------ACAGATTATCT 5628

Qy        729 aIleHis----------------------------------------------------- 731
               :::
Db       5629 GGTTGAAATTAACAAAATTTACTTTGCATGGATGATGTTGAATTATCGCTTAATGTTCC 5688

Qy        732 ------------CysIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGlnIl 747
                          |||:::      ||||||   :::|||                ||
Db       5689 AGAGCTCAACACTGCTATTTACGAAGACTTCTCTTTTCAGGAAGACTCTCTGAAGAATAT 5748

Qy        747 ePheGluProLeuHisLys------------------------------------------ 753
              |   :::   |||   |||
Db       5749 CAAAGACCAACTGGACAAACTTGGAGAGCAGATTGCAGTCATTCATGAAAAACAGCCAGA 5808

Qy        754 -----SerLeuAspProSerAsnLeuGlu-----------------HisLeuIle---- 764
                   ||||||  |||   :::|||:::                 ||||||:::
Db       5809 TGTCATCCTTGAAGCCTCTGGACCTGAAGCCATTCAGATCAGAGATACACTTACTCAGCT 5868

Qy        765 -------------ThrProLeuVal---------ThrIleGlyHisIleAlaLeuLeuA 777
                           |||   |||:::          ||||||   :::   |||
Db       5869 GAATGCAAAATGGGACAGAATTAATAGAATGTACAGTGATCGGAAAGGTTGTTTTG---- 5924

Qy        777 laProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspL 797
                            ||||||:::       ::::::   :::
Db       5925 --------------ACAGGGCAATGGAAGAATGGAGACAGTTCCATTGTGACC------- 5963

Qy        797 euLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrp---ValProAspG 816
                ||||||                 ::::::          |||   :::|||
Db       5964 --TTAATGACCTCACACAGTGGATAACAGAGGCTGAAGAATTACTGGTTGATACCTGTGC 6021

Qy        816 luGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpL 836
                :::|||:::      |||     |||               :::||||||
Db       6022 TCCAGGTGGCAGCCTGGACTTAGAGAAAGCCAGGATACATCAGCAGGAACTTGAGGTGGG 6081

Qy        836 euLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrT 856
                :::        ::::::|||      |||       :::
Db       6082 C-ATC---------AGCAGCCACCAGCCCAGTTTTGCAGCA------------------ 6112

Qy        856 hrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetS 876
                  |||:::        ||||||         |||   |||    |||:::|||:::   |||  |
Db       6113 -----CTAAACCGAACTGGGGATGGGATTGTGCAG---AAACTCTCCCAGGCAGATGGAA 6164

Qy        876 erArgLeu---------------------------ArgLeuAlaAlaGlySerAlaIleV 887
              ||  |||                            |||    |||    |||  |
Db       6165 GCTTCTTGAAAGAAAAACTGGCAGGTTTAAACCAACGCTGGGATGCAATTGTTGCAGAAG 6224

Qy        887 alLysLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGluGlnTyrGln---- 905
                |||||     |||                   |||         :::   :::|||:::
Db       6225 TGAAGGATAGGCAGCCAAGGCTAAAAGGAGAAAGTAAGCAGGTGATGAAGTACAGGCATC 6284

Qy        906 -------------LeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnV 921
                           :::|||   |||             :::          |||   |||
```

```
Db       6285 AGCTAGATGAGATTATCTGTTGGTTAACAAAGGCTGAGCATGCTATGCAAAAGAGATCA- 6343

Qy        921 alPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetA 941
                           :::|||    :::    |||     |||||
Db       6344 -----ACCACCGAATTGGGAGAAAACCTGCAAGAATTAAGA------------------ 6379

Qy        941 laIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnC 961
                                    |||       :::|||         ||||||   :::
Db       6380 ----------------------GACTTAACTCAAGAAATGGAAGTACATGCTGAAAAA- 6415

Qy        961 ysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerG 981
                    :::|||   :::|||      |||      |||:::           :::|||:
Db       6416 -----CTCAAATGGCTGAAT------AGAACTGAATTGGAGATG---------CTTTCAG 6455

Qy        981 luLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaH 1001
              :::|||   |||||||||
Db       6456 ATAAAAGTCTGAGTTTA------------------------------------------ 6472

Qy       1001 isAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspVal--------- 1017
                    |||:::      |||:::         |||    |||:::    |||
Db       6473 -----CCTGAAAGGGATAAAATTTCA------GAAAGCTTAAGGACTGTAAATATGACAT 6521

Qy       1018 ----------------------------------LysGluCysLeuTrpPheValLeuG 1026
                                                |||||||||:::
Db       6522 GGAATAAGATTTGCAGAGAGGTGCCTACCACCCTGAAGGAATGCATC------------ 6568

Qy       1026 luIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluA 1046
                                                         :::          :
Db       6569 -----------------------------------------------CAGGAGCCCAGTT 6581

Qy       1046 snIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuT 1066
              :::::   ||||||:::    |||        |||:::    |||:::
Db       6582 CTGTTTCACAGACAAGGATTGCTGCTCATCCTAATGTCCAAAAGGTG------------ 6628

Qy       1066 yrThrValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSerLeuG 1086
                    :::         |||   :::|||    :::           |||      :::
Db       6629 --GTGCTAGTATCATCTGCGTCAGATATTCCTGTTCAGTCTCATCGTACTTCGGAAATTT 6686

Qy       1086 luSerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsnPheS 1106
                    |||    |||                               ||||||       :
Db       6687 CAATTCCTGCTGAT-----------------------------CTTGATAAAACTATAA 6716

Qy       1106 erAsnThrLysAsnTyrLeu---------ProProGluMetLysSer---PhePheThrP 1122
              ::          ::::::|||                 :::||||||         |||
Db       6717 CAGAACTAGCCGACTGGCTGGTATTAATCGACCAGATGCTGAAGTCCAACATTGTCACTG 6776

Qy       1122 roGlyLys-----------ProLysThrThrAsnValLeuGlyAlaValAsnLysProL 1138
                |||                 ||||||    :::     :::                |
Db       6777 TTGGGGATGTAGAAGAGATCAATAAGACCGTTTCCCGAATGAAAATTACAAAGGCTGACT 6836

Qy       1138 euSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnA 1158
              ||                |||                 :::         :::    :::    |||
Db       6837 TAGAACAGCGCCATCCTCAGCTGGATTATGTTTTTACATTGGCACAGAATTTGAAAAATA 6896

Qy       1158 laSerSerSerSerAsnProSerSerPro---------------GlyArgIleLysGlyA 1173
                 :::||||||||||:::       :::                  |||:::|||      :
Db       6897 AAGCTTCCAGTTCAGATATGAGAACAGCAATTACAGAAAAATTGGAAAGGGTCAAGAACC 6956

Qy       1173 rgLeuAspSerSer-----------GluMetAspHisSerGluAsnGluAspTyrThrM 1189
              ::    |||    :::           |||:::         :::    ||||||       :
Db       6957 AGTGGGATGGCACCCAGCATGGCGTTGAGCTAAGACAGCAGCAGCTTGAGGACATGATTA 7016

Qy       1189 etSerSerProLeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgS 1209
              ::    |||               :::      |||:::::::::::::     :::
Db       7017 TTGACAGTCTT---------CAGTGGGATGACCATAGGGAGGAGACTGAAGAACTGATGA 7067

Qy       1209 erGluLeuGluLysPro-----------ArgGlyArgLysLysThrProValThrGluG 1225
                 :::    |||                 :::       ::::::    |||:::|||:::|
Db       7068 GAAAATATGAGGCTCGACTCTATATTCTTCAGCAAGCCCGACGGGATCCACTCACCAAAC 7127

Qy       1225 ln--------------------GluGluLysLeuGlyMetAspAsp------------- 1233
              ||                    ::::::||||||      |||
Db       7128 AAATTTCTGATAACCAAATACTGCTTCAAGAACTGGGTCCTGGAGATGGTATCGTCATGG 7187

Qy       1234 -------------LeuThrLysLeuValGlnGluGlnLysProLysGlySerGlnArgS 1249
                           |||    ||||||::::::|||              ||||||       :
Db       7188 CGTTCGATAACGTCCTGCAGAAACTCCTGGAGGAATAT---------GGGAGTGATGACA 7238

Qy       1249 erArgLysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrpProGluGluLysA 1269
              :::|||             |||    :::|||          :::          |||
Db       7239 CAAGGAATGTGAAAGAAACC---ACAGAGTACTTAAAAACATCATGG---------ATCA 7286

Qy       1269 rgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsnSerProProLysLys-Gly 1288
                 ||||||:::    |||                 |||    ||||||:::           ::: |||
Db       7287 ATCTCAAACAAAGTATT---------GCTGACAGACAGAACGCCTTGGAGGCTGAGTGGA 7337

Qy       1289 LysArgGlyArgProProLysProLeuGlyGlyGlyThrProLysGluGluProThrMet 1308
                 |||    ||||||             :::
Db       7338 GGACGGTGCAGGCCTCTCGCAGAGATC----------------------------TGG 7367

Qy       1309 LysThrSerLysLysGlySerLysLysLys 1318
                 |||||||||       ||||||||||||:::
Db       7368 AAAACTTCCTGAAGTGGATCCAAGAAGCAG 7397

RESULT 9
US-09-134-001C-322
; Sequence 322, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 322
;   LENGTH: 30549
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-322

Alignment Scores:
Pred. No.:                5.9e-06        Length:        30549
Score:                    190.00         Matches:       276
Percent Similarity:       35.29%         Conservative:  235
Best Local Similarity:    19.06%         Mismatches:    523
Query Match:              2.64%          Indels:        414
DB:                       4              Gaps:          67

US-09-512-581B-2 (1-1391) x US-09-134-001C-322 (1-30549)

Qy         23 SerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValValLysThrPhe 42
              :::|||:::|||                      :::       :::|||       :::|||
Db      26830 AATGATGAAATTATGAATACTAATTCTATTGCACAATTGTCTAGAATCGTGAATCAAGCA 26889

Qy         43 MetAspMetAspGlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAlaLeuHis 62
                 |||::::::             :::       ::::::             |||    |||
Db      26890 TTTGATTTAAATGATGCAATGAAATCTTTAAGAGATGAACTTAATAATCAAGCTTTTCCT 26949

Qy         63 LeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuValAlaCys 82
              :::    :::                :::       :::|||::::::
Db      26950 GTCCAAGCAAGCTCAAATTATATAAATTCAGATGAAGATTTAAAACAACAATTTGACCAT 27009

Qy         83 CysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLys 102
                 |||::::::          ::::::    |||    |||                :::|||
Db      27010 GCTTTAAGTAATGCTCGAAAAGTTCTTGCAAAAGAAAATGGTAAAAATTTAGATGAAAAA 27069

Qy        103 LeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeu-----------Glu 118
                                               |||::::::||||||             |||
Db      27070 ----------------------------CAAATTCAGGGACTCAAACAAGTGATTGAG 27099
```

```
Qy      119 AspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpVal 138
            |||||||||
Db    27100 GATACTAAAGAT------------------------------------------------ 27111

Qy      139 LysSerTyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyr 158
            :::      |||            |||      :::               |||
Db    27112 ---GCTTTAAATGGTATCCAACGTTTATCAAAAGCTAAAGCTAAAGCAATTCAATACGTA 27168

Qy      159 ArgThrLeuPheSerValIleAsnAsnGly------HisAsnGlnLysValHisMetHis 176
            ::::::|||   |||   ||||||:::         |||      :::   ::::::|||
Db    27169 CAATCTTTA---TCTTATATCAATGATGCACAGCGTCATATTGCTGAAAATAATATTCAC 27225

Qy      177 MetValAspLeuMetSerSerIle-----------------IleCysGluGlyAspThr 190
                  |||   :::||||||:::                        :::   |||
Db    27226 AACTCTGATGATTTATCATCTTTAGCAAATACATTATCTAAAGCTAGTGATTTAGATAAT 27285

Qy      191 ValSerGlnGluLeuLeuAspThrVal---------------------------LeuVal 201
                  ::::::|||   ||||||:::                              |||
Db    27286 GCAATGAAAGACTTACGAGATACTATAGAAAGTAATTCAACTTCTGTTCCAAATAGTGTG 27345

Qy      202 AsnLeuValProAlaHisLysAsnLeuAsn---------LysGlnAlaTyrAspLeuAla 218
            |||   :::   |||   |||||||||                  :::|||      |||
Db    27346 AATTATATTAATGCTGATAAGAATTTACAAATTGAATTTGATGAGGCGCTACAACAAGCA 27405

Qy      219 LysAlaLeuLeuLysArgThr--------------------------------------- 225
               |||         :::|||
Db    27406 AGTGCAACAAGTTCTAAAACTTCAGAAAATCCAGCAACGATTGAAGAAGTATTAGGTCTT 27465

Qy      226 AlaGlnAlaIle---------------------------GluProTyrIleThrThrPhe 236
            :::|||||||||                           |||      :::   |||
Db    27466 AGTCAAGCCATTTACGATACAAAAAATGCATTAAATGGTGAACAACGACTTGCAACTGAG 27525

Qy      237 PheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPhe 256
            ::::::      |||   |||   |||            ||||||::::::
Db    27526 AAGAGCAAAGATCTAAAATTAATAAAA---GGATTAAAAGATTTAAATAAAGCACAACTT 27582

Qy      257 AspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGln 276
            :::   :::               ::::::|||         |||:::   |||   |||
Db    27583 GAAGATGTCACAAAC------AAGGTAAATTCAGCAAATACTTTAACAGAGTTATCTCAG 27636

Qy      277 Leu---GluPheLysLeuLysSerAsnAspAsnGluGluArgLeuGln---ValValLys 294
            |||   :::      |||:::   ||||||      :::   |||:::      :::|||
Db    27637 CTCACTCAATCAACGTTAGAATTAAACGATAAAATGAAATTATTGAGAGATAAGCTTAAA 27696

Qy      295 LeuLeuAlaLysMetPheGlyAlaLys------------------AspSerGluLeuAla 308
               |||               |||                     |||      |||
Db    27697 ACTTTAGTAAATCCTGTTAAAGCAAGTTTAAATTATAGAAACGCTGATTATAATTTAAAA 27756

Qy      309 SerGln---AsnLysProLeuTrpGlnCys------TyrLeuGlyArgPheAsnAspIle 325
            |||      ||||||   |||   :::            |||   :::   :::
Db    27757 CGTCAATTTAACAAAGCTTTAAAAGAAGCTAAAGGCGTATTAAATAAAAAATAGCGGTACA 27816

Qy      326 HisValProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisPro 345
            :::|||   |||                           |||   |||:::
Db    27817 AATGTCAATATC--------------AATGACATTCAACAT---CTTTTAACACAAATA 27858

Qy      346 AspLeuAlaLysAsp---LeuThrGluTyrLeuLysValArgSerHisAspProGluGlu 364
            |||   |||||||||   |||            :::::::::   |||      :::
Db    27859 GATAATGCTAAAGACCAATTAAATGGTGAACGACGTCTAAAAGAACATCAACAAAAATCT 27918

Qy      365 AlaIle-----------ArgHisAspValIleValSerIleValThrAla--------- 377
               :::                  |||::::::   :::         |||
Db    27919 GAAGTATTTATTATTAAAGAATTAGATATACTTAATAATGCTCAAAAAGCTGCAATAATT 27978

Qy      378 ------------AlaLysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheVal 393
                        |||   |||||||||   ::::::|||      :::   |||   :::
Db    27979 AATCAGATTAGAGCGTCTAAAGACATTAAAATAATTAATCAAATCGTTGATAATGCAATA 28038

Qy      394 ArgGluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMetMetGlyLeuAla 413
                                                   :::||||||   ||||||
Db    28039 GAA---------------------------TTAAATGATGCTATGCAAGGTTTA--- 28065

Qy      414 GlnIleTyrLysLysTyrAlaLeuGln---SerAlaAlaGlyLysAspAlaAlaLysGln 432
                     |||::::::      |||   :::|||      ||||||
Db    28066 ---------AAAGAACATGTAGCTCAATTAACAGCAACTACAAAAGAC----------- 28104

Qy      433 IleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArg 452
                                       :::|||   |||   |||:::   :::|||
Db    28105 --------------------------AACATTGAATATTTAAATGCTGATGAAGACCAT 28137

Qy      453 LeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAsnLeuGlu--------- 469
            |||                     ||||||      ||||||
Db    28138 AAATTA-----------------CAATATGATTACGCTATCAACTTAGCGAATAATGTT 28179

Qy      470 -----------------ThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThr 483
                              |||   :::            :::      :::
Db    28180 CTTGACAAAGAAAACGGTACAAATAAAGACGCTAATATCATAATTGGAATGATTCAAAAC 28239

Qy      484 LeuAsp---------------LeuAsnAlaValLysAlaLeuAsnGlu---MetTrpLys 497
            :::|||               ||||||   ::::::   |||   :::         |||
Db    28240 ATGGATGATGCTAGAGCACTTCTAAATGGAATTGAAAGACTTAAAGATGCTCAAACAAAA 28299

Qy      498 CysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLys 517
                  |||   ::::::      :::|||      ||||||   |||:::
Db    28300 GCACATAATGACATTAAAGATACGCTCAAACGTCAACTTGATGAAATTGAACACGCTAAT 28359

Qy      518 ThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuPro 537
                  ::::::   ||||||   :::|||   ||||||
Db    28360 GCAACATCAAATTCTAAAGCT---CAAGCTAAACAAATGGTA----------------- 28398

Qy      538 AspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGlu 557
                                                                  :::|||
Db    28399 ---------------------------------------------------AATGAG 28404

Qy      558 LysIleArgLysGlnLeuGluValLeuValSerProThrCysSer------CysLysGln 575
            :::   ||||||   |||      :::         |||   :::            |||
Db    28405 GAAGCTAGAAAAGCGCTTTCTAATATTAATGACGCAACATCAAATGATTTAGTTAATCAA 28464

Qy      576 Ala------GluGly-------------------CysValArgGluIleThrLysLys 586
            |||      ||||||                              |||:::   |||
Db    28465 GCAAAAGATGAAGGGCAATCTGCAATTGAACACATACATGCAGATGAATTACCTAAAGCA 28524

Qy      587 LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGlu 606
                  :::      |||      :::      :::|||   |||      |||:::
Db    28525 AAACTAGATGCTAATCAAATGATTGACCAAAAAGTTGAAGATATAAATCACTTAATTAGT 28584

Qy      607 ArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsn 626
            :::            ::::::      |||      :::   ||||||   |||:::|||
Db    28585 CAA------AATCCAAACTTATCAAATGAAGAAAAAAATAAACTAATATCTCAAATTAAT 28638

Qy      627 LysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIle 646
            |||   ::::::|||      ::::::   :::   :::            |||
Db    28639 AAGTTAGTAAATGGAATTAAGAATGAAATTCAACAAGCTATAAACAAACAACAAATAGAA 28698

Qy      647 ArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSer 666
               |||      :::|||   :::|||:::      |||      :::
Db    28699 AATGCTACAACAAAACTAGATGAAGTCATTGAAACTACTAAAAAATTAATTATCGCCAAA 28758

Qy      667 AlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGlu 686
            ||||||
Db    28759 GCAGAA------------------------------------------------------ 28764

Qy      687 AlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHisIle 706
               |||   |||:::   |||            |||      :::
Db    28765 ---GCTAAACAAATGATAAAAGAGTTATCACAAAAGAAACGAGAT--------------- 28806

Qy      707 ArgSerAlaLeuLeuProValLeuHisHisLysSerLysLysGlyProProArgGlnAla 726
                                 :::::::::   :::         |||   ::::::|||
Db    28807 -----------------GCAATAAATAACAACACTGATTTAACACCTTCTCAAAAGGCA 28848

Qy      727 LysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGln 746
                        |||   :::      |||      :::::::::
Db    28849 ------------CATGCTTTAGCAGATATTGATAAAACAGAAAAAGATGCACTTCAACAT 28896

Qy      747 IlePheGluProLeuHisLysSerLeuAspPro------SerAsnLeuGluHisLeuIle 764
            |||                  |||:::|||         :::|||   ||||||
```

```
Db  28897 ATCGAAAAT------TCTAATTCAATTGATGATATCAATAACAATAAAGAGCATGCATTT 28950

Qy    765 ThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln----------- 780
                      |||:::   ||||||    :::         ||||||
Db  28951 AAT---------ACTTTAGCTCATATCATTATTTGGGATACTGATCAGCAACCATTAGTT 29001

Qy    781 PheAlaAlaPro------------TrpLysSerTrpValAlaThrPheIleVal------ 794
          |||      |||                     :::   |||   :::   :::|||
Db  29002 TTTGAAGTACCTGAATTGAGCCTTCAAAATGCTCTAGTAACAAGTGAGGTGGTTGTTCAC 29061

Qy    795 ---------------------------------LysAspLeuLeuMetAsnAspArgLeu 803
                                            |||   :::   :::   |||   |||
Db  29062 AGAGATGAAACTATTTCATTAGAATCTATAAAAAAAAAAATGACTTTAACTGATGAACTT 29121

Qy    804 ProGlyLysLysThrThrLysLeuTrpValProAsp---GluGluValSerProGluThr 822
                   |||         :::   :::|||:::   :::::|||:::
Db  29122 ---------AAAGTCAATATTGTTTCATTACCGAACACTGATAAAGTAGCTGATCACCTA 29172

Qy    823 MetValLysIleGlnAlaIle--------------------------------------- 829
                |||::::::   |||
Db  29173 ACCGCTAAAGTTAAGGTTATTTTAGCTGATGGCTCATATGTCACTGTAAATGTTCCAGTC 29232

Qy    830 LysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSer 849
          |||::::::   :::   |||         |||   :::      |||:::
Db  29233 AAAGTTGTAGAAAAAGAATTACAAATAGCTAAAAAGGATGCTATAAAAACAATTGATGTT 29292

Qy    850 ThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGlu---GlnGly 868
             ::::::               :::   |||:::   :::||||||      |||
Db  29293 CTGGTAAAACAAAAAATCAAAGATATAGATTCTAATAACGAATTAACGTCTACTCAACGT 29352

Qy    869 LysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleValLys 888
          :::   :::|||   ::::::   ||||||:::                ||||||   |||
Db  29353 GAAGATGCAAAAGCTGAAATTGAAAGATTGAAA---------AAGCAAGCCATCGATAAA 29403

Qy    889 LeuAlaGlnGluProCysTyrHisGluIleIleThr------------------------ 900
          :::                     :::|||   |||
Db  29404 GTGAATCATTCAAAATCGATTAAAGATATTGAAACAGTAAAACGAACTGATTTTGAAGAA 29463

Qy    901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsn------------------------ 912
          ::::::|||:::                  :::|||
Db  29464 ATAGATCAGTTTGATCCTAAACGCTTTACGCTAAATAAAGCTAAAAAGGATATCATTACT 29523

Qy    913 AspGluCysTyrGlnValArgGlnValPheAlaGln---LysLeuHisLysGlyLeuSer 931
          |||         |||::::::      |||      :::   :::      |||||||||:::
Db  29524 GATGTTAATACTCAAATCCAAAATGGTTTCAAAGAAATTGAAACAATAAAAGGTTTAACT 29583

Qy    932 ArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProVal 951
                                        :::   ||||||            :::
Db  29584 TCTAATGAAAAAACTCAGTTTGATAAACAATTAACTGCACTACAAAAGAATTTTTAGAA 29643

Qy    952 LysGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGlu 971
          |||         ||||||                  ||||||:::            :::::::::|||
Db  29644 AAAGTCGAGCATGCTCATAAT---------TTAGTAGAATTAAATCAATTACAACAAGAG 29694

Qy    972 Tyr---------------------LeuLysGlnHisAlaAlaValSerGluLysLeuLeu 984
          :::                     |||   |||         :::      ||||||
Db  29695 TTTAATAATAGATATAAACATATTTAAACCAAGCACATTTACTAGGTGAAAAA------ 29748

Qy    985 SerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAsp 1004
                                              |||:::            |||
Db  29749 -----------------------------------CATATAGCAGAACATAAATTAGGA 29772

Qy   1005 TyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheVal 1024
          ||||||   |||         :::|||:::   :::   :::         :::|||:::
Db  29773 TATGTTGTAGTAAACAAAACTCAGCAAATACTAAATAATCAATCTGCTTCTTACTTTATA 29832

Qy   1025 LeuGlu-------------------------------------------------IleLeu 1028
             :::                                                   |||
Db  29833 AAACAATGGGCACTTGATAGAATTAAACAAATTCAACTAGAAACGATGAATTCAATTCGT 29892

Qy   1029 MetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLys 1048
             |||            :::   |||         :::   :::   :::|||   |||
Db  29893 GGTGCGCATACCGTACAAGATGTACACAAAGCATTATTACAAGGTATAGAGCAAATCTTG 29952

Qy   1049 GlnThrLysAspAla--------------------------------------------- 1053
          :::         :::
Db  29953 AAAGTAAATGTAAGTATTATAAATCAATCTTTCAACGATTCCTTGCATAACTTTAATTAT 30012

Qy   1054 ---GlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAla 1072
                         ||||||::::::   ||||||                ||||||||
Db  30013 CTTCATTCAAAATTTGATGCTAGATTAAGAGAAAAG--------------GATGTTGCA 30057

Qy   1073 MetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProVal 1092
             :::|||:::   :::::::::                  |||:::            |||
Db  30058 AACCATATCGTACAAACTGAAACATTCAAAGAAGTTCTAAAAGGA------ACGGGTGTT 30111

Qy   1093 LeuProAlaArgPhe------PheThrGlnProAspLysAsnPheSerAsnThrLysAsn 1110
             |||   :::            ||||||      |||         |||      :::
Db  30112 GAACCAGGTAAAATCAACAAAGAAACACAGCAACCAAAACTTCATAAGAATGATAATGAT 30171

Qy   1111 TyrLeuProProGluMetLysSerPhePhe------ThrProGlyLysProLysThrThr 1128
             |||      :::            |||      |||   |||            |||
Db  30172 AGCCTATTCAAACATTTAGTTGATAATTTCGGCAAAACTGTAGGTGTTATTACATTAACT 30231

Qy   1129 AsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLys 1148
             :::|||   :::            |||:::         |||:::   :::   :::
Db  30232 GGTTTACTTTCTAGTTTCTGGTTAGTTTTGGCTAAAAGACGTAAAAAAGAAGAAGAAGAA 30291

Qy   1149 SerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGly 1168
                   ::::::
Db  30292 AAACAATCGATAAAA--------------------------------------------- 30306

Qy   1169 ArgIleLysGlyArgLeuAspSerSerGluMetAspHisSerGluAsnGluAspTyrThr 1188
                                                 :::   ::::::|||
Db  30307 -------------------------------------AATCATCACAAAGATATTCGT 30327

Qy   1189 MetSerSerProLeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArg 1208
          :::|||                     :::||||||   |||
Db  30328 CTTTCAGAT------------------ACTGATAAAATAGATCCAATTGTAATAACTAAG 30369

Qy   1209 SerGluLeuGluLysProArgGlyArgLysLysThrProValThrGluGlnGluGluLys 1228
             :::::::::|||                                     ||||||:::
Db  30370 CGTAAAATAGATAAA------------------------------------GAAGAACAA 30393

Qy   1229 LeuGlyMetAspAspLeuThrLysLeu------ValGlnGluGlnLysProLysGlySer 1246
          :::      ||||||         :::         :::   :::|||         |||
Db  30394 ATTCAAAACGATGACAAACATTCAATTCCAGTTGCTAAACATAAGAAATCTAAAGAAAAG 30453

Qy   1247 GlnArgSerArgLysArgGlyHisThr------------------AlaSerGluSerAsp 1260
          |||   |||   :::      |||:::                           ||||||
Db  30454 CAATTGAGTGAAGAGGATATTCATTCAATCCCCGTCGTTAAGCGTAAACAAAACAGTGAT 30513

Qy   1261 GluGlnGlnTrpProGluGluLys 1268
             |||::::::      |||:::
Db  30514 AACCAAAGATACAAAACAGAAGAA 30537
```

RESULT 10
US-08-139-937-12
; Sequence 12, Application US/08139937
; Patent No. 5821070
; GENERAL INFORMATION:
; APPLICANT: LEE, WEN-HWA
; APPLICANT: SHAN, BEI
; TITLE OF INVENTION: CELLULAR GENES ENCODING
; TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/139,937
;      FILING DATE:  20-OCT-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/979,156
;      FILING DATE:  20-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAMPBELL, CATHRYN
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-CJ 9370
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-535-9001
;      TELEFAX:  619-535-8949
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4868 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-139-937-12

Alignment Scores:
Pred. No.:                2.52e-07      Length:         4868
Score:                    189.50        Matches:        277
Percent Similarity:       35.50%        Conservative:   247
Best Local Similarity:    18.77%        Mismatches:     608
Query Match:              2.63%         Indels:         345
DB:                       1             Gaps:           61

US-09-512-581B-2 (1-1391) x US-08-139-937-12 (1-4868)

Qy         19 ValLysGluIleSerAspLysIle-----------------------SerLysGluGlu 30
              :::||||||:::       |||:::                       :::|||   |||
Db        156 ATGAAAGAATTAGACTCAAAACTCCATTTACAGGAGGTACAACTAATGACCAAAATTGAA 215

Qy         31 MetValArgArgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGlu 50
                             |||:::   :::|||                   :::   |||
Db        216 GCATGCATAGAATTGGAAAAAATAGTTGGGGAACTTAAGAAAGAAAACTCAGATTTAAGT 275

Qy         51 GluGluLysGluLeuTyr------------------------------------------ 56
              |||:::   |||   :::
Db        276 GAAAAATGGAATATTTTTCTTGTGATCACCAGGAGTTACTCCAGAGAGTAGAAACTTCT 335

Qy         57 ------LeuAsnLeuAlaLeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLys 74
                    ||||||   |||   :::   :::|||         |||         :::
Db        336 GAAGGCCTCAATTCTGATTTAGAAATGCATGCAGAT---------AAATCATCACGTGAA 386

Qy         75 AspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGlu 94
              |||:::         ||||||
Db        387 GATATTGGAGATAATGTGGCC--------------------------------------- 407

Qy         95 AlaProTyrThrSerProAspLysLeuLysAspIlePheMetPheIleThrArgGlnLeu 114
                                |||      |||:::   |||:::   :::      :::|||
Db        408 ---------AAGGTGAATGACAGCTGGAAGGAGAGATTTCTTGATGTGGAAAATGAGCTG 458

Qy        115 LysGlyLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsn 134
                    :::         |||:::         :::         ||||||      :::
Db        459 AGTAGGATCAGATCGGAGAAAGCTAGCATTGAGCATGAAGCCCTCTACCTGGAGGCTGAC 518

Qy        135 IleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSerAsn------Glu 152
              :::         |||::::::      :::|||   |||   ::::::   |||      :::
Db        519 TTAGAGGTAGTTCAAACAGAGAAGCTATGTTTAGAAAAAGACAATGAAAATAAGCAGAAG 578

Qy        153 IlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAsnGlyHisAsn------ 170
              :::      |||         |||   ||||||:::   :::      |||
Db        579 GTTATTGTCTGCCTTGAAGAAGAACTC---TCAGTGGTCACAAGTGAGAGAAACCAGCTT 635

Qy        170 ------------------------------------------------------------ 170

Db        636 CGTGGAGAATTAGATACTATGTCAAAAAAAACCACGGCACTGGATCAGTTGTCTGAAAAA 695

Qy        171 ---------------GlnLysValHisMetHisMetValAspLeuMetSerSerIle--- 184
                             |||::::::      |||         :::   :::      |||
Db        696 ATGAAGGAGAAAACACAAGAGCTTGAGTCTCATCAAAGTGAGTGTCTCCATTGCATTCAG 755

Qy        185 IleCysGluGlyAspThrValSerGln---GluLeuLeuAspThrValLeuValAsnLeu 203
              :::   |||   :::         :::   |||||||||   |||:::      ::::::
Db        756 GTGGCAGAGGCAGAGGTGAAGGAAAAGACGGAACTCCTTCAGACTTTGTCCTCTGATGTG 815

Qy        204 ValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeu------ 221
                          |||   |||               |||      :::   |||
Db        816 -----------AGTGAGCTGTTAAAAGACAAAACTCATCTCCAGGAAAAGCTGCAGAGT 863

Qy        222 LeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPheAsnGlnValLeu 241
              |||::::::   :::||||||:::                           ||||||:::
Db        864 TTGGAAAAGGACTCACAGGCACTGTCTTTGACAAAATGTGAGCTGGAAAACCAAATTGCA 923

Qy        242 MetLeuGlyLys------------------ThrSerIleSerAspLeuSerGluHisVal 255
                 |||   |||                                |||         :::
Db        924 CAACTGAATAAAGAGAAAGAATTGCTTGTCAAGGAATCTGAAAGCCTGCAGGCCAGACTG 983

Qy        256 PheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuPro 275
                 :::         |||      |||:::            |||      ::::::::
Db        984 AGTGAATCAGATTATGAAAAGCTGAATGTCTCCAAGGCCTTGGAGGCCGCACTGGTGGAG 1043

Qy        276 GlnLeuGluPheLysLeuLysSerAsnAspAsnGluGlu---------------ArgLeu 290
              :::   ||||||   |||:::   :::         :::|||                  |||
Db       1044 AAAGGTGAGTTCGCATTGAGGCTGAGCTCAACACAGGAGGAAGTGCATCAGCTGAGAAGA 1103

Qy        291 GlnValValLysLeuLeuAlaLysMetPheGly-AlaLysAspSerGluLeuAlaSerGl 310
                 :::   ||||||         ::::::       |||   |||         |||||
Db       1104 GGCATCGAGAAACTGAGAGTTCGCATTGAGGCCGATGAAAAGAAGCAGCTGCACATCGCA 1163

Qy        310 nAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIleAr 330
              |                 :::         ||||||      ||||||:::         ::
Db       1164 GAGAAAC------TGAAAGAACGCGAGCGGGAGAATG------ATTCACTTAAGGTAAAA 1211

Qy        330 gLeuGluCysValLysPheAlaSerHisCys-------------------------Le 341
              :|||      :::|||         |||                              ||
Db       1212 GTTGAGAACCTTGAAAGGGAATTGCAGATGTCAGAAGAAAACCAGGAGCTAGTGATTCTT 1271

Qy        341 uMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHisAs 361
              ||||         |||   :::   |||         |||   ||||||   :::      ::
Db       1272 GATG------CCGAGAATTCCAAAGCAGAA-GTAGAGACTCTAAAAACACAAATAGAAGA 1324

Qy        361 pProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLysAs 381
              :            :::::::::         :::         :::         ::::::|||::
Db       1325 GATGGCCAGAAGCCTGAAAGTTTTGAATTAGACCTTGTCACGTTAAGGTCTGAAAAAGA 1384

Qy        381 pIleLeuLeuValAsn--------------------------------AspHisLeuLe 390
              :   |||                                         |||   |||||
Db       1385 AAATCTGACAAAACAAATACAAGAAAAACAAGGTCAGTTGTCAGAACTAGACAAGTTACT 1444

Qy        390 uAsnPheValArgGluArgThrLeuAspLysArgTrp---------ArgValArgLysGl 407
              |:::         :::               :::|||            ::::::::::::||
Db       1445 CTCTTCATTTAAAAGTCTGTTAGAAGAAAAGGAGCAAGCAGAGATACAGATCAAAGAAGA 1504

Qy        407 uAlaMetMetGlyLeuAlaGlnIleTyrLys------LysTyrAlaLeuGlnSerAlaAl 425
              |:::         :::      :::               |||         ::::::   ||
Db       1505 ATCTAAAACTGCAGTGGAGATGCTTCAGAATCAGTTAAAGGAGCTAAATGAGGCAGTAGC 1564

Qy        425 aGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTy 445
              |                  ::::::|||               ::::::   |||
Db       1565 AGCCTTGTGTGGTGACCAAGAAATTATGAAGGCCACAGAACAGAGTCTAGACCCACCAAT 1624

Qy        445 rGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValPr 465
                 :::                  |||         :::|||::::::   |||:::
Db       1625 AGAGGAAGAGCATCAGCTGAGAAATAGCATTGAAAAGCTGAGAGCCCGC----------- 1673

Qy        465 oHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAs 485
                       ||||||      |||:::   :::
Db       1674 -------CTAGAAGCTGATGAAAAGAAGCAG----------------------------- 1697
```

```
Qy        485 pLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGl 505
              |||       ::::::  |||  |||               :::||||||:::   ::
Db       1698 -CTCTGTGTCTTACAACAACTGAAGGAAAGTGAGCATCATGCAGATTTACTTAAGGGTAG 1756

Qy        505 nValLysAspLeu---------LeuAspLeuIleLysGlnProLysThrAspAlaSerVa 522
              :|||::::::|||         |||::::::  :::       :::     |||:::::
Db       1757 AGTGGAGAACCTTGAAAGAGAGCTAGAGATAGCCAGGACAAACCAAGAGCATGCAGCTCT 1816

Qy        522 lLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAl 542
              ::::|||     ||||||    |||  |||                     |||
Db       1817 TGAGGCAGAGAATTCCAAAGGAGAGGTAGAGACCCTAAAGCAAAAATAGAAGGGATGAC 1876

Qy        542 aGlnAspPhe-------------------------------------MetLysLysPh 549
               |||                                             :::
Db       1877 CCAAAGTCTGAGAGGTCTGGAATTAGATGTTGTTACTATAAGGTCAGAAAAAGAAAATCT 1936

Qy        549 eThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerPr 569
               |||    |||:::  ::::::::         :::||||||::::::   |||
Db       1937 GACAAATGAATTACAAAAAGAGCAAGAGCGAATATCTGAATTAGAAATAATAAATTCA-- 1994

Qy        569 oThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAs 589
                                  |||    :::::::|||     ::::::
Db       1995 ----------------TCATTTGAAAATATTTTGCAAGAAAAAGAGCAAGAGAAAGTACA 2038

Qy        589 nProLysGlnProThrAsnProPheLeuGluMet------------IleLysPheLeuLe 605
                 |||:::   ::::::     :::||||||             :::|||   |||
Db       2039 GATGAAAGAAAAATCAAGCACTGCCATGGAGATGCTTCAAACACAATTAAAAGAGCTCAA 2098

Qy        605 uGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnVa 625
               ||||||:::|||   :::|||   |||   |||:::    |||       |||:::
Db       2099 TGAGAGAGTGGCAGCCCTGCATAATGACCAAGAAGCCTGTAAGGCC------AAAGAGCA 2152

Qy        625 lAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAl 645
               |||   |||                        :::   |||        ::::::||
Db       2153 GAATCTTAGT------------------AGTCAAGTAGAGTGTCTTGAACTTGAGAAGGC 2194

Qy        645 a---IleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPh 664
              |    :::     ||||||:::     |||              :::   |||
Db       2195 TCAGTTGCTACAAGGCCTTGATGAGGCCAAA------------AATAATTATATTGTTTT 2242

Qy        664 eHisSerAla---------------GluThrPheGluSerLeuLeuAlaCysLeuLysMe 679
                 |||:::                 :::      |||            |||:::
Db       2243 GCAATCTTCAGTGAATGGCCTCATTCAAGAAGTAGAAGATGGCAAGCAGAAACTGGAGAA 2302

Qy        679 tAspAspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIl 699
                 ||||||:::::::::            ||||||
Db       2303 GAAGGATGAAGAAATCAGTAGACTGAAAAATCAAATT----------------------- 2339

Qy        699 eGluGluAspPheProHisIleArgSerAlaLeuLeuProVal------LeuHisHis-- 716
                  :::|||       :::   |||   |||   |||         |||      |||
Db       2340 ----CAAGACCAAGAGCAGCTTGTCTCTAAACTGTCCCAGGTGGAAGGAGAGCACCAACT 2395

Qy        717 ----LysSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAl 735
                  |||   :::           |||               :::
Db       2396 TTGGAAGGAGCAAAACTTAGAACTGAGAAATCTGACAGTGGAATTGGAGCAGAAGATCCA 2455

Qy        735 aIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLe 755
               :::         ||||||                    |||   |||      |||
Db       2456 AGTGCTACAATCCAAAAATGCCTCTTTGCAGGACACATTAGAAGTGCTGCAGAGTTCTTA 2515

Qy        755 uAspProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLe 775
                          |||||||||:::                       :::   ||
Db       2516 C------AAGAATCTAGAGAAT---------------------------GAGCTTGAATT 2542

Qy        775 uLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLy 795
              |            |||:::                          :::|||:::
Db       2543 GACAAAAATGGACAAAATG--------------------------TCCTTTGTT----- 2570

Qy        795 sAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAs 815
                                            :::|||      |||:::             ::
Db       2571 ---------------------------------GAAAAAGTAAACAAAATGACTGCAAAGGA 2599

Qy        815 pGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTr 835
              :   |||:::          |||                |||          :::
Db       2600 AACTGAGCTGCAGAGGGAAATGCATGAGATGGCACAGAAAACAGCAGAGCTGCAAGAAGA 2659

Qy        835 pLeuLeuGlyMetLysAsn-----------------------------AsnHisSerLy 845
               |||   |||   ||||||                                     ||
Db       2660 ACTCAGTGGAGAGAAAAATAGGCTAGCTGGAGAGTTGCAGTTACTGTTGGAAGAAATAAA 2719

Qy        845 sSerGlyThrSerThrLeuArgLeuLeuThr----------------------------- 855
               ||||        |||:::   ||||||
Db       2720 GAGCAGCAAAGATCAATTGAAGGAGCTCACACTAGAAAATAGTGAATTGAAGAAGAGCCT 2779

Qy        856 -ThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMe 875
                 :::|||   |||      :::     ::::::||||||:::   :::   :::::
Db       2780 AGATTGCATGCACAAAGAC---CAGGTGGAAAAGGAAGGGAAAGTG---AGAGAGGAAAT 2833

Qy        875 tSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTy 895
               ::::      :::|||            |||      |||
Db       2834 AGCTGAATATCAGCTACGGCTTCATGAAGCTGAAAAGAAA-------------------- 2873

Qy        895 rHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCy 915
               |||:::   :::   |||:::                                ::::::
Db       2874 -CACCAGGCTTTGCTTTTGGAC--------------------------ACAAACAAACA 2905

Qy        915 sTyrGlnValArg---GlnValPheAlaGlnLysLeu---------HisLysGlyLeuSe 931
               |||:::|||      |||   :::   :::||||||            :::   |||||
Db       2906 GTATGAAGTAGAAATCCAGACATACCGAGAGAAATTGACTTCTAAAGAAGAATGTCTCAG 2965

Qy        931 rArgLeuArgLeuProLeuGluTyrMetAlaIleCys------------------AlaLe 945
              |     :::|||   ::::::   :::                               :::||
Db       2966 TTCACAGAAGCTGGAGATAGACCTTTTAAAGTCTAGTAAAGAAGAGCTCAATAATTCATT 3025

Qy        945 uCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAs 965
              |   |||        :::   |||            :::            :::|||
Db       3026 GAAAGCTACTACTCAGATTTTGGAAGAATTGAAGAAAACCAAGATGGACAATCTAAAATA 3085

Qy        965 nIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSe 985
               :::|||              |||||||:::              ::::::      :::
Db       3086 TGTAAAT------------CAGTTGAAGAAGGAAAATGAACGTGCCCAGGGGAAAATGAA 3133

Qy        985 rLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTy 1005
               ||||||
Db       3134 GTTGTTG----------------------------------------------------- 3140

Qy       1005 rValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValLe 1025
                    ::::::      :::||||||::::::      ||||||   |||        ||
Db       3141 -------ATCAAATCCTGTAAACAGCTGGAAGAGGAAAAGGAGATACTGCAGAAAGAACT 3193

Qy       1025 uGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArg----------- 1041
              |     |||   |||      |||                  :::
Db       3194 CTCTCAACTTCAAGCTGCACAGGAGAAGCAGAAAACAGGTACTGTTATGGATACCAAGGT 3253

Qy       1042 ----LysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLy 1060
                  ::::::         ||||||:::   |||:::                       ::
Db       3254 CGATGAATTAACAACTGAGATCAAAGAACTGAAAGAAACTCTTGAAGAAAAAACCAAGGA 3313

Qy       1060 sMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSe 1080
              :     :::|||   |||         |||               ::::::::::::|||
Db       3314 GGCAGATGAATACTTGGATAAGTACTGT------------TCCTTGCTTATAAGCCATGA 3361

Qy       1080 rThrThrTyrSerLeuGluSerProLysAsp------ProValLeuProAlaArgPhePh 1098
                             ||||||   |||:::                         |||
Db       3362 A---------AAGTTAGAGAAAGCTAAAGAGATGTTAGAGACACAAGTGGCCCATCTGTG 3412

Qy       1098 eThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSe 1118
               :::|||      |||      |||                  |||   |||   :::
Db       3413 TTCACAGCAATCTAAACAAGATTCCCGAGGGTCTCCTTTGCTAGGTCCAGTTGTT----- 3467

Qy       1118 rPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLe 1138
                                  |||   :::   :::                  |||   ||
Db       3468 ------------------CCAGGACCATCTCCAATCCCTTCTGTTACTGAAAAGAGGTT 3508

Qy       1138 uSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAl 1158
```

```
                    ||||||        :::::      |||      |||        :::
Db       3509 ATCATCTGGCCAAAATAAAGCTTCAGGCAAGAGGCAAAGATCCAGTGGAATATGGGAGAA 3568

Qy       1158 aSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGl 1178
                                   |||:::::|||               :::        :::
Db       3569 TGGTGGAGGACCAACACCTGCTACCCCAGAGACCTTTTCTAAAAAAAGCAAGAAAGCAGT 3628

Qy       1178 uMetAsp---------HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGl 1195
                |||               :::|||:::   :::   |||        |||
Db       3629 CATGAGTGGTATTCACCCTGCAGAAGACACGGAAGGTACTGAGTTTGAGCCAGAGGGACT 3688

Qy       1195 yLysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysProAr 1215
                 :::       |||:::      :::|||:::          :::
Db       3689 TCCAGAAGTTGTAAAGAAAGGGTTTGCTGACATCCCGACAGGAAAGACTAGCCCATATAT 3748

Qy       1215 gGlyArgLysLysThrProValThrGluGlnGlu---------------GluLysLeuGl 1230
                  |||:::   |||       |||                        ||||||
Db       3749 CCTGCGAAGAACAACCATGGCAACTGGGAGCAGGCCCGGCCTGGCTGCACACAAGTTACC 3808

Qy       1230 yMetAspAspLeuThr------------LysLeuValGlnGluGlnLysPro---LysGl 1245
                :::     ||||||               |||   :::     ||||||      ||
Db       3809 CCTATCCCCACTGACTGTCCCCAAACAAAATCTTGCAGAGTCCTCCAAACCAACAGCTGG 3868

Qy       1245 ySerGlnArgSerArgLys---ArgGlyHisThrAlaSerGluSerAspGluGlnGlnTr 1264
               |          ||||||:::|||                |||        |||
Db       3869 TGGCAGCAGATCACAAAAGGTGAAAGTTGCTCAGCGGAGCCCAGTAGATTCAGGCACCAT 3928

Qy       1264 pProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsn----- 1282
                     |||        |||        :::    |||     |||:::::::
Db       3929 CCTCCGAGAACCCACCACGAAATCCGTCCCAGTCAATAATCTTCCTGAGAGAAGTCCGAC 3988

Qy       1283 -SerProProLysLysGly------------LysArgGlyArgProProLysProLeuGl 1298
                     |||::::::|||               :::|||         |||   |||
Db       3989 TGACAGCCCCAGAGAGGGCCTGAGGGTCAAGCGCCGGCGACTTGTCCCCAGCCCCAAAGC 4048

Qy       1298 yGly-GlyThrProLysGluGluProThrMetLysThrSerLys 1312
                |||     :::|||:::            |||:::::::::|||:::
Db       4049 TGGACTGGAGTCCAAGGGCAGTGAGAACTGTAAGGTCCAGTGAA 4092
```

RESULT 11
PCT-US93-11310-12
; Sequence 12, Application PC/TUS9311310
; GENERAL INFORMATION:
; APPLICANT: BOARD OF REGENTS OF THE UNIVERSITY OF TEXAS SYSTEM
; TITLE OF INVENTION: CELLULAR GENES ENCODING
; TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11310
; FILING DATE: 19-NOV-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHRYN
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: FP-CJ 9790
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4868 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
PCT-US93-11310-12

```
Alignment Scores:
Pred. No.:              2.52e-07     Length:        4868
Score:                  189.50       Matches:       277
Percent Similarity:     35.50%       Conservative:  247
Best Local Similarity:  18.77%       Mismatches:    608
Query Match:            2.63%        Indels:        345
DB:                     5            Gaps:          61
```

US-09-512-581B-2 (1-1391) x PCT-US93-11310-12 (1-4868)

```
Qy         19 ValLysGluIleSerAspLysIle-----------------------SerLysGluGlu 30
              :::||||||:::      |||:::                       :::|||   |||
Db        156 ATGAAAGAATTAGACTCAAAACTCCATTTACAGGAGGTACAACTAATGACCAAAATTGAA 215

Qy         31 MetValArgArgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGlu 50
                               |||:::   :::|||                 :::    |||
Db        216 GCATGCATAGAATTGGAAAAAATAGTTGGGGAACTTAAGAAAGAAAACTCAGATTTAAGT 275

Qy         51 GluGluLysGluLeuTyr------------------------------------------ 56
              |||:::   |||   :::
Db        276 GAAAAATTGGAATATTTTCTTGTGATCACCAGGAGTTACTCCAGAGAGTAGAAACTTCT 335

Qy         57 ------LeuAsnLeuAlaLeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLys 74
                    ||||||      |||   :::   :::|||          |||         :::
Db        336 GAAGGCCTCAATTCTGATTTAGAAATGCATGCAGAT---------AAATCATCACGTGAA 386

Qy         75 AspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGlu 94
              |||:::         ||||||
Db        387 GATATTGGAGATAATGTGGCC--------------------------------------- 407

Qy         95 AlaProTyrThrSerProAspLysLeuLysAspIlePheMetPheIleThrArgGlnLeu 114
                           |||      |||:::   |||:::    :::         :::|||
Db        408 ---------AAGGTGAATGACAGCTGGAAGGAGAGATTTCTTGATGTGGAAAATGAGCTG 458

Qy        115 LysGlyLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsn 134
                 :::          |||:::          :::          ||||||       :::
Db        459 AGTAGGATCAGATCGGAGAAAGCTAGCATTGAGCATGAAGCCCTCTACCTGGAGGCTGAC 518

Qy        135 IleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSerAsn------Glu 152
              :::      |||::::::         :::|||   |||   ::::::   |||   :::
Db        519 TTAGAGGTAGTTCAAACAGAGAAGCTATGTTTAGAAAAAGACAATGAAAATAAGCAGAAG 578

Qy        153 IlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAsnGlyHisAsn------ 170
              :::         |||         |||   ||||||:::   :::      |||
Db        579 GTTATTGTCTGCCTTGAAGAAGAACTC---TCAGTGGTCACAAGTGAGAGAAACCAGCTT 635

Qy        170 ------------------------------------------------------------ 170

Db        636 CGTGGAGAATTAGATACTATGTCAAAAAAAACCACGGCACTGGATCAGTTGTCTGAAAAA 695

Qy        171 --------------GlnLysValHisMetHisMetValAspLeuMetSerSerIle--- 184
                            |||::::::      |||         :::   :::   |||
Db        696 ATGAAGGAGAAAACACAAGAGCTTGAGTCTCATCAAAGTGAGTGTCTCCATTGCATTCAG 755

Qy        185 IleCysGluGlyAspThrValSerGln---GluLeuLeuAspThrValLeuValAsnLeu 203
              :::   |||   :::            :::   |||||||||   |||:::      ::::::
Db        756 GTGGCAGAGGCAGAGGTGAAGGAAAAGACGGAACTCCTTCAGACTTTGTCCTCTGATGTG 815

Qy        204 ValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeu------ 221
                                   |||   |||               |||   :::   |||
Db        816 ------------AGTGAGCTGTTAAAAGACAAAACTCATCTCCAGGAAAAGCTGCAGAGT 863

Qy        222 LeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnValLeu 241
              |||::::::      :::||||||:::                   ||||||:::
Db        864 TTGGAAAAGGACTCACAGGCACTGTCTTTGACAAAATGTGAGCTGGAAAACCAAATTGCA 923

Qy        242 MetLeuGlyLys-----------------ThrSerIleSerAspLeuSerGluHisVal 255
```

```
                 |||  |||                                 |||        :::
Db        924 CAACTGAATAAAGAGAAAGAATTGCTTGTCAAGGAATCTGAAAGCCTGCAGGCCAGACTG 983

Qy        256 PheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuPro 275
               :::         |||    |||:::       |||      :::::::::
Db        984 AGTGAATCAGATTATGAAAAGCTGAATGTCTCCAAGGCCTTGGAGGCCGCACTGGTGGAG 1043

Qy        276 GlnLeuGluPheLysLeuLysSerAsnAspAsnGluGlu---------------ArgLeu 290
              :::   ||||||  |||:::   :::    :::|||                 |||
Db       1044 AAAGGTGAGTTCGCATTGAGGCTGAGCTCAACACAGGAGGAAGTGCATCAGCTGAGAAGA 1103

Qy        291 GlnValValLysLeuLeuAlaLysMetPheGly-AlaLysAspSerGluLeuAlaSerGl 310
              :::   ||||||        ::::::           |||   |||         |||||
Db       1104 GGCATCGAGAAACTGAGAGTTCGCATTGAGGCCGATGAAAAGAAGCAGCTGCACATCGCA 1163

Qy        310 nAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIleAr 330
              |          :::        ||||||       ||||||:::          ::
Db       1164 GAGAAAC------TGAAAGAACGCGAGCGGGAGAATG------ATTCACTTAAGGTAAAA 1211

Qy        330 gLeuGluCysValLysPheAlaSerHisCys--------------------------Le 341
              :|||       :::|||      |||                               ||
Db       1212 GTTGAGAACCTTGAAAGGGAATTGCAGATGTCAGAAGAAAACCAGGAGCTAGTGATTCTT 1271

Qy        341 uMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHisAs 361
              ||||      |||  :::  |||       |||  ||||||  :::           ::
Db       1272 GATG------CCGAGAATTCCAAAGCAGAA-GTAGAGACTCTAAAAACACAAATAGAAGA 1324

Qy        361 pProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLysAs 381
              :          :::::::::        :::           :::          :::::|||::
Db       1325 GATGGCCAGAAGCCTGAAAGTTTTGAATTAGACCTTGTCACGTTAAGGTCTGAAAAAGA 1384

Qy        381 pIleLeuLeuValAsn--------------------------------AspHisLeuLe 390
              :  |||                                        |||  |||||
Db       1385 AAATCTGACAAAACAAATACAAGAAAAACAAGGTCAGTTGTCAGAACTAGACAAGTTACT 1444

Qy        390 uAsnPheValArgGluArgThrLeuAspLysArgTrp---------ArgValArgLysGl 407
              |:::       :::         :::|||          ::::::::::::||
Db       1445 CTCTTCATTTAAAAGTCTGTTAGAAGAAAAGGAGCAAGCAGAGATACAGATCAAAGAAGA 1504

Qy        407 uAlaMetMetGlyLeuAlaGlnIleTyrLys------LysTyrAlaLeuGlnSerAlaAl 425
              |:::     :::       :::          |||         ::::::    ||
Db       1505 ATCTAAAACTGCAGTGGAGATGCTTCAGAATCAGTTAAAGGAGCTAAATGAGGCAGTAGC 1564

Qy        425 aGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTy 445
              |                 ::::::|||           ::::::  |||
Db       1565 AGCCTTGTGTGGTGACCAAGAAATTATGAAGGCCACAGAACAGAGTCTAGACCCACCAAT 1624

Qy        445 rGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValPr 465
               :::              |||      :::|||::::::   |||:::
Db       1625 AGAGGAAGAGCATCAGCTGAGAAATAGCATTGAAAAGCTGAGAGCCCGC----------- 1673

Qy        465 oHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAs 485
                      ||||||     |||:::  :::
Db       1674 -------CTAGAAGCTGATGAAAAGAAGCAG----------------------------- 1697

Qy        485 pLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGl 505
               |||       ::::::  |||  |||               :::||||||:::   ::
Db       1698 -CTCTGTGTCTTACAACAACTGAAGGAAAGTGAGCATCATGCAGATTTACTTAAGGGTAG 1756

Qy        505 nValLysAspLeu---------LeuAspLeuIleLysGlnProLysThrAspAlaSerVa 522
              :|||:::::|||          |||::::::   :::        :::     |||:::::
Db       1757 AGTGGAGAACCTTGAAAGAGAGCTAGAGATAGCCAGGACAAACCAAGAGCATGCAGCTCT 1816

Qy        522 lLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGlyLysAl 542
              ::::|||      ||||||       |||   |||                  |||
Db       1817 TGAGGCAGAGAATTCCAAAGGAGAGGTAGAGACCCTAAAAGCAAAAATAGAAGGGATGAC 1876

Qy        542 aGlnAspPhe--------------------------------------MetLysLysPh 549
               |||                                             :::
Db       1877 CCAAAGTCTGAGAGGTCTGGAATTAGATGTTGTTACTATAAGGTCAGAAAAAGAAAATCT 1936

Qy        549 eThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerPr 569
               |||       |||::: . :::::::::         :::||||||:::::   |||
Db       1937 GACAAATGAATTACAAAAAGAGCAAGAGCGAATATCTGAATTAGAAATAATAAATTCA-- 1994

Qy        569 oThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAs 589
                                      |||    :::::::|||        ::::::
Db       1995 ----------------TCATTTGAAAATATTTTGCAAGAAAAAGAGCAAGAGAAAGTACA 2038

Qy        589 nProLysGlnProThrAsnProPheLeuGluMet-----------IleLysPheLeuLe 605
                |||:::   ::::::        :::||||||              :::|||  |||
Db       2039 GATGAAAGAAAAATCAAGCACTGCCATGGAGATGCTTCAAACACAATTAAAAGAGCTCAA 2098

Qy        605 uGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnVa 625
               ||||||:::|||   :::|||   |||   |||:::      |||      |||:::
Db       2099 TGAGAGAGTGGCAGCCCTGCATAATGACCAAGAAGCCTGTAAGGCC------AAAGAGCA 2152

Qy        625 lAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAl 645
                |||   |||                          :::   |||   :::   ::::::||
Db       2153 GAATCTTAGT-----------------AGTCAAGTAGAGTGTCTTGAACTTGAGAAGGC 2194

Qy        645 a---IleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPh 664
              |   :::     ||||||:::       |||                 :::   |||
Db       2195 TCAGTTGCTACAAGGCCTTGATGAGGCCAAA------------AATAATTATATTGTTTT 2242

Qy        664 eHisSerAla---------------GluThrPheGluSerLeuLeuAlaCysLeuLysMe 679
                 |||:::                 :::       |||        |||:::
Db       2243 GCAATCTTCAGTGAATGGCCTCATTCAAGAAGTAGAAGATGGCAAGCAGAAACTGGAGAA 2302

Qy        679 tAspAspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIl 699
                ||||||:::::::::                ||||||
Db       2303 GAAGGATGAAGAAATCAGTAGACTGAAAAATCAAATT----------------------- 2339

Qy        699 eGluGluAspPheProHisIleArgSerAlaLeuLeuProVal------LeuHisHis-- 716
                :::|||          :::  |||         |||              |||
Db       2340 ----CAAGACCAAGAGCAGCTTGTCTCTAAACTGTCCCAGGTGGAAGGAGAGCACCAACT 2395

Qy        717 ----LysSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAl 735
                  |||  :::           |||                 :::
Db       2396 TTGGAAGGAGCAAAACTTAGAACTGAGAAATCTGACAGTGGAATTGGAGCAGAAGATCCA 2455

Qy        735 aIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLe 755
               :::     ||||||                       |||  |||      |||
Db       2456 AGTGCTACAATCCAAAAAATGCCTCTTTGCAGGACACATTAGAAGTGCTGCAGAGTTCTTA 2515

Qy        755 uAspProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLe 775
                    |||||||||:::                             :::   ||
Db       2516 C------AAGAATCTAGAGAAT---------------------------GAGCTTGAATT 2542

Qy        775 uLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLy 795
              |           |||:::                             ::::|||:::
Db       2543 GACAAAAATGGACAAAATG---------------------------TCCTTTGTT----- 2570

Qy        795 sAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAs 815
                                           :::|||     |||:::           ::
Db       2571 ---------------------------------GAAAAAGTAAACAAAATGACTGCAAAGGA 2599

Qy        815 pGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTr 835
              :   |||:::         |||          |||            :::
Db       2600 AACTGAGCTGCAGAGGGAAATGCATGAGATGGCACAGAAAACAGCAGAGCTGCAAGAAGA 2659

Qy        835 pLeuLeuGlyMetLysAsn-----------------------------AsnHisSerLy 845
                |||   |||   ||||||                                   ||
Db       2660 ACTCAGTGGAGAGAAAAATAGGCTAGCTGGAGAGTTGCAGTTACTGTTGGAAGAAATAAA 2719

Qy        845 sSerGlyThrSerThrLeuArgLeuLeuThr---------------------------- 855
              ||||          |||:::   ||||||
Db       2720 GAGCAGCAAAGATCAATTGAAGGAGCTCACACTAGAAAATAGTGAATTGAAGAAGAGCCT 2779

Qy        856 -ThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMe 875
                    :::|||   |||      :::   ::::::||||||:::   :::   :::::
Db       2780 AGATTGCATGCACAAAGAC---CAGGTGGAAAAGGAAGGGAAAGTG---AGAGAGGAAAT 2833

Qy        875 tSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTy 895
              ::::      :::|||             |||      |||
Db       2834 AGCTGAATATCAGCTACGGCTTCATGAAGCTGAAAAGAAA------------------- 2873
```

```
Qy        895 rHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCy 915
               |||:::  :::  |||:::                                  ::::::
Db       2874 -CACCAGGCTTTGCTTTTGGAC--------------------------ACAAACAAACA 2905

Qy        915 sTyrGlnValArg---GlnValPheAlaGlnLysLeu---------HisLysGlyLeuSe 931
               |||:::|||      |||   :::  :::||||||            :::  |||||
Db       2906 GTATGAAGTAGAAATCCAGACATACCGAGAGAAATTGACTTCTAAAGAAGAATGTCTCAG 2965

Qy        931 rArgLeuArgLeuProLeuGluTyrMetAlaIleCys------------------AlaLe 945
              |     :::|||  ::::::   :::                              :::||
Db       2966 TTCACAGAAGCTGGAGATAGACCTTTTAAAGTCTAGTAAAGAAGAGCTCAATAATTCATT 3025

Qy        945 uCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAs 965
              |   |||             :::   |||             :::          :::|||
Db       3026 GAAAGCTACTACTCAGATTTTGGAAGAATTGAAGAAAACCAAGATGGACAATCTAAAATA 3085

Qy        965 nIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSe 985
               :::|||              ||||||:::         ::::::       :::
Db       3086 TGTAAAT-----------CAGTTGAAGAAGGAAAATGAACGTGCCCAGGGGAAAATGAA 3133

Qy        985 rLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTy 1005
               ||||||
Db       3134 GTTGTTG----------------------------------------------------- 3140

Qy       1005 rValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValLe 1025
                    ::::::       :::|||||||::::::    ||||||    |||          ||
Db       3141 -------ATCAAATCCTGTAAACAGCTGGAAGAGGAAAAGGAGATACTGCAGAAAGAACT 3193

Qy       1025 uGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArg----------- 1041
              |       |||  |||       |||                 :::
Db       3194 CTCTCAACTTCAAGCTGCACAGGAGAAGCAGAAAACAGGTACTGTTATGGATACCAAGGT 3253

Qy       1042 ----LysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLy 1060
                   ::::::            |||||||:::   |||:::                    ::
Db       3254 CGATGAATTAACAACTGAGATCAAAGAACTGAAAGAAACTCTTGAAGAAAAAACCAAGGA 3313

Qy       1060 sMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSe 1080
              :   :::|||  |||          |||             ::::::::::::|||
Db       3314 GGCAGATGAATACTTGGATAAGTACTGT-----------TCCTTGCTTATAAGCCATGA 3361

Qy       1080 rThrThrTyrSerLeuGluSerProLysAsp------ProValLeuProAlaArgPhePh 1098
                           ||||||       |||:::                     |||
Db       3362 A---------AAGTTAGAGAAAGCTAAAGAGATGTTAGAGACACAAGTGGCCCATCTGTG 3412

Qy       1098 eThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSe 1118
               :::|||       |||       |||                 |||   |||    :::
Db       3413 TTCACAGCAATCTAAACAAGATTCCCGAGGGTCTCCTTTGCTAGGTCCAGTTGTT----- 3467

Qy       1118 rPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLe 1138
                                   |||      :::   :::                    |||  ||
Db       3468 ------------------CCAGGACCATCTCCAATCCCTTCTGTTACTGAAAAGAGGTT 3508

Qy       1138 uSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAl 1158
               |||||||          ::::::      |||       |||           :::
Db       3509 ATCATCTGGCCAAAATAAAGCTTCAGGCAAGAGGCAAAGATCCAGTGGAATATGGGAGAA 3568

Qy       1158 aSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGl 1178
                                |||::::::|||                  :::          :::
Db       3569 TGGTGGAGGACCAACACCTGCTACCCCAGAGACCTTTTCTAAAAAAAGCAAGAAAGCAGT 3628

Qy       1178 uMetAsp---------HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGl 1195
                 |||                :::|||:::   :::    |||          |||
Db       3629 CATGAGTGGTATTCACCCTGCAGAAGACACGGAAGGTACTGAGTTTGAGCCAGAGGGACT 3688

Qy       1195 yLysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysProAr 1215
                 :::          |||:::           :::|||:::            :::
Db       3689 TCCAGAAGTTGTAAAGAAAGGGTTTGCTGACATCCCGACAGGAAAGACTAGCCCATATAT 3748

Qy       1215 gGlyArgLysLysThrProValThrGluGlnGlu---------------GluLysLeuGl 1230
                 |||:::  |||       |||                                 ||||||
Db       3749 CCTGCGAAGAACAACCATGGCAACTGGGAGCAGGCCCGGCCTGGCTGCACACAAGTTACC 3808

Qy       1230 yMetAspAspLeuThr------------LysLeuValGlnGluGlnLysPro---LysGl 1245
                :::      ||||||                |||   :::          ||||||    ||
Db       3809 CCTATCCCCACTGACTGTCCCCAAACAAAATCTTGCAGAGTCCTCCAAACCAACAGCTGG 3868

Qy       1245 ySerGlnArgSerArgLys---ArgGlyHisThrAlaSerGluSerAspGluGlnGlnTr 1264
                |        ||||||:::|||   :::                 |||       |||
Db       3869 TGGCAGCAGATCACAAAAGGTGAAAGTTGCTCAGCGGAGCCCAGTAGATTCAGGCACCAT 3928

Qy       1264 pProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsn----- 1282
                        |||         |||          :::    |||        |||::::::
Db       3929 CCTCCGAGAACCCACCACGAAATCCGTCCCAGTCAATAATCTTCCTGAGAGAAGTCCGAC 3988

Qy       1283 -SerProProLysLysGly------------LysArgGlyArgProProLysProLeuGl 1298
                         |||::::::|||                ::::|||        |||    |||
Db       3989 TGACAGCCCCAGAGAGGGCCTGAGGGTCAAGCGCCGGCGACTTGTCCCCAGCCCCAAAGC 4048

Qy       1298 yGly-GlyThrProLysGluGluProThrMetLysThrSerLys 1312
                 |||    :::|||:::            |||:::::::::|||:::
Db       4049 TGGACTGGAGTCCAAGGGCAGTGAGAACTGTAAGGTCCAGTGAA 4092
```

RESULT 12
US-08-836-022A-10/c
; Sequence 10, Application US/08836022A
; Patent No. 6001557
; GENERAL INFORMATION:
;   APPLICANT: Trustees of the University of Pennsylvania
;   APPLICANT: Wilson, James M.
;   APPLICANT: Fisher, Krishna J.
;   APPLICANT: Chen, Shu-Jen
;   APPLICANT: Weitzman, Matthew
;   TITLE OF INVENTION: Improved Adenovirus Virus and
;   NUMBER OF SEQUENCES: 10
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Howson and Howson
;   STREET: Spring House Corporate Cntr, P O Box 457
;   CITY: Spring House
;   STATE: Pennsylvania
;   COUNTRY: USA
;   ZIP: 19477
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/836,022A
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/331,381
;   FILING DATE: 28-OCT-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Bak, Mary E.
;   REGISTRATION NUMBER: 31,215
;   REFERENCE/DOCKET NUMBER: GNVPN.008PCT
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 215-540-9200
;   TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 10:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 19307 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: unknown
;   MOLECULE TYPE: cDNA
US-08-836-022A-10

Alignment Scores:

Pred. No.:	8.87e-06	Length:	19307
Score:	184.50	Matches:	241
Percent Similarity:	33.38%	Conservative:	213
Best Local Similarity:	17.72%	Mismatches:	496
Query Match:	2.56%	Indels:	411

```
DB:                    3            Gaps:          50

US-09-512-581B-2 (1-1391) x US-08-836-022A-10 (1-19307)

Qy       206 AlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThr 225
             |||   :::   ||||||   :::      :::      :::   |||   :::|||
Db     12049 GCTTCAGAACAACTGAACAGCCGGTGGACAGAATTCTGCCAATTGCTGAGTGAGAGAGTT 11990

Qy       226 Ala---GlnAlaIleGluProTyrIleThrThrPhePheAsnGlnValLeuMetLeuGly 244
                      :::            |||   ||||||:::||||||:::      |||
Db     11989 AACTGGCTAGAGTATCAAACCAACATCATTACCTTTTATAATCAGCTACAACAATTGGAA 11930

Qy       245 ------------------------------LysThrSerIleSerAspLeuSerGlu--- 253
                                           ||||||      :::   |||||||||
Db     11929 CAGATGACAACTACTGCCGAAAACTTGTTGAAAACCCAGTCTACCACCCTATCAGAGCCA 11870

Qy       254 HisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerVal 273
                               |||::::::         :::   :::   |||   ::::::
Db     11869 ACAGCAATTAAAAGCCAGTTAAAAATTTGTAAGGATGAAGTCAACAGATTGTCAGCTCTT 11810

Qy       274 LeuProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValVal 293
                ||||||:::|||                        :::   ::::::|||   :::
Db     11809 CAGCCTCAAATTGAG------------------------CAATTAAAAATTCAGAGTCTA 11774

Qy       294 LysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysPro 313
             :::|||      |||      |||                                 |||
Db     11773 CAACTGAAAGAAAAGGGACAGGGG---------------------------------CCA 11747

Qy       314 LeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCys 333
             ::::::                                       :::      :::
Db     11746 ATGTTT---------------------------------------CTGGATGCAGACTTT 11726

Qy       334 ValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGlu 353
             |||   |||   :::|||         ||||||
Db     11725 GTGGCCTTTACTAATCAT------TTTAACCAC--------------------------- 11699

Qy       354 TyrLeuLysValArgSerHisAspProGluGluAlaIleArgHisAspValIleValSer 373
                                                                |||
Db     11698 ---------------------------------------------------ATCTTTGAT 11690

Qy       374 IleValThrAlaAlaLysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheVal 393
                |||   |||   :::|||::::::      :::   |||   |||
Db     11689 GGTGTGAGGGCCAAAGAGAAAGAGCTACAGACAATTTTGACACTTTACCACCAATGCGC 11630

Qy       394 ArgGluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMetMetGlyLeuAla 413
                :::   |||:::               :::|||
Db     11629 TATCAGGAGACAATGAGTAGC---------ATCAGG------------------------ 11603

Qy       414 GlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIle 433

Db     11603 ------------------------------------------------------------ 11603

Qy       434 AlaTrpIleLys------AspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAsp 451
                ||||||:::      :::      |||   :::   |||   :::      :::
Db     11602 ACGTGGATCCAGCAGTCAGAAAGCAAACTCTCTGTACCTTATCTTAGTGTTACTGAATAT 11543

Qy       452 ArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrThr 471
                ::::::   ||||||:::                        :::|||:::
Db     11542 GAAATAATGGAGGAGAGACTCGGGAAATTACAGGCTCTGCAAAGTTCTTTGAAA------ 11489

Qy       472 GluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAla 491
                |||:::            ||||||      |||:::         |||   ||||||
Db     11488 GAGCAACAAAATGGCTTCAACTATCTGAGTGACACTGTGAAGGAGATGGCCAAGAAAGCA 11429

Qy       492 LeuAsnGluMet------------------------------------------TrpLys 497
                :::|||:::                                          ||||||
Db     11428 CCTTCAGAAATATGCCAGAAATATCTGTCAGAATTTGAAGAGATTGAGGGGCACTGGAAG 11369

Qy       498 ---------------------------CysGlnAsnLeuLeuArgHisGlnValLysAsp 508
                                        ||||||   |||      |||
Db     11368 AAACTTTCCTCCCAGTTGGTGGAAAGCTGCCAAAAGCTAGAAGAACAT------------ 11321

Qy       509 LeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLys 528
                      :::   |||      |||            :::|||   :::      |||
Db     11320 ---------ATGAATAAACTTCGAAAATTTCAGAATCACATAAAAACCTTA---CAGAAA 11273

Qy       529 ValMetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLys 548
                |||                              :::         |||:::|||:::
Db     11272 TGGATG---------------------------GCTGAAGTTGATGTTTTCCTGAAAGAG 11240

Qy       549 PheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSer 568
                         |||   |||   |||   ::::::|||||||||:::
Db     11239 GAATGGCCTGCCCTGGGGGATGCTGAAATCCTGAAAAAACAGCTCAAA------------ 11192

Qy       569 ProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGly 588
                         |||:::         |||   ::::::   |||         |||
Db     11191 ---------CAATGCAGACTTTAGTTGGTGATATTCAAACAATTCAGCCCAGTTTAAAT 11141

Qy       589 AsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIle 608
             :::               |||         :::   ||||||         |||
Db     11140 AGT------------GTTAATGAAGGTGGGCAGAAGATAAAGAGTGAAGCTGAACTTGAG 11093

Qy       609 AlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSer 628
                         ::::::||||||               ::::::::::::|||
Db     11092 TTTGCATCCAGACTGGAGACAGAA---------------CTTAGAGAGCTTAACACTCAG 11048

Qy       629 IleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArgAla 648
                |||                     :::               :::|||::::::|||
Db     11047 TGGGATCACATATGC---------CGCCAGGTCTACACCAGAAAGGAAGCCTTAAAGGCA 10997

Qy       649 GlyLeuGlu------------LeuLeuLysValLeuSerPheThrHisProIleSerPhe 664
             ||||||:::            |||   |||   ||||||      |||
Db     10996 GGTTTGGATAAAACCGTAAGCCTCCAAAAAGATCTATCAGAGATGCATGAGTGGATGACA 10937

Qy       665 HisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGlu------ 682
                :::   |||   :::                     |||      ||||||
Db     10936 CAAGCTGAAGAAGAATATCTAGAGAGAGATTTTGAATATAAAACTCCAGATGAATTACAG 10877

Qy       683 ---------------------LysValAlaGluAlaAlaLeuGln--------------- 690
                                  :::      |||   |||||||||
Db     10876 ACTGCTGTTGAAGAAATGAAGAGAGCTAAAGAAGAGGCACTACAAAAAGAAACTAAAGTG 10817

Qy       691 ------IlePheLysAsnThrGlySerLysIleGluGluAspPheProHisIleArgSer 708
                   :::   :::         |||   |||            |||      :::
Db     10816 AAACTCCTTACTGAGACTGTAAATAGTGTAATAGCTCACGCTCCACCCTCAGCACAAGAG 10757

Qy       709 AlaLeuLeuProValLeuHisHisLysSerLys--------------------------- 719
             ||||||   :::
Db     10756 GCCTTAAAAAAGGAACTTGAAACTCTGACCACCAACTACCAATGGCTGTGCACCAGGCTG 10697

Qy       720 LysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSer 739
                |||      :::      :::            |||   |||   :::   |||
Db     10696 AATGGAAAATGCAAAACTTTGGAAGAAGTTTGGGCATGTTGGCATGAGTTATTGTCATAT 10637

Qy       740 LysGlu------------------------------------------------------ 741
                |||
Db     10636 TTAGAGAAAGCAAACAAGTGGCTCAATGAAGTAGAATTGAAACTTAAAACCATGGAAAAT 10577

Qy       742 ---------------ThrGlnPheAlaGlnIlePheGluProLeu--------------- 751
                               :::      ::::::   |||   |||
Db     10576 GTTCCTGCAGGACCTGAGGAAATCACTGAAGTGCTAGAATCTCTTGAAAATCTGATGCAT 10517

Qy       752 HisLysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGly 771
             |||         :::|||:::   :::         |||      |||         |||
Db     10516 CATTCAGAGGAGAACCCAAATCAGATTCGTCTATTGGCACAGACTCTTACAGATGGAGGA 10457

Qy       772 HisIleAlaLeuLeuAlaProAspGln---------PheAlaAlaProTrpLysSerTrp 788
                :::      |||      ::::::         |||   :::   |||:::
Db     10456 GTCATGGATGAACTGATCAATGAGGAGCTTGAGACGTTTAATTCTCGTTGGAGGGAACTA 10397

Qy       789 ValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThr 808
                         :::   |||      |||:::         :::         :::
Db     10396 CATGAAGAGGCTGTGAGGAAACAAAAGTTGCTTGAACAGAGTATCCAGTCTGCCCAGGAA 10337

Qy       809 ThrLys---------------------------LeuTrpValProAspGluGluValSer 819
                :::                           |||         |||::::::::::::
```

```
Db     10336 ATTGAAAAGTCCTTGCACTTAATTCAGGAGTCGCTTGAATTCATTGACAAGCAGTTGGCA 10277

Qy       820 ProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMet 839
                                   |||:::    |||    :::|||
Db     10276 GCTTATATCACTGACAAGGTGGATGCAGCTCAAATGCCTCAGGAA-----------GCC 10229

Qy       840 LysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHis 859
             :::          |||          :::       :::   |||               |||
Db     10228 CAGAAAATCCAATCAGATTTGACAAGTCATGAGATAAGTTTAGAAGAAATGAAGAAACAT 10169

Qy       860 SerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArg 879
             :::   |||              |||       :::|||:::   |||:::::::::   :::
Db     10168 AACCAGGGGAAGGATGCCAACCAAAGGGTTCTTTCACAAATTGATGTTGCACAGAAAAAA 10109

Qy       880 LeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGlu------ 897
             |||       |||          :::|||   |||:::|||          |||
Db     10108 TTACAAGATGTCTCCATGAAATTTCGATTATTCCAAAAACCAGCCAATTTTGAACAACGT 10049

Qy       898 -----------------IleIleThrLeuGluGlnTyrGlnLeuCysAlaLeu------ 909
                               |||:::          :::      |||   ||||||
Db     10048 CTAGAGGAAAGTAAGATGATTTTAGATGAAGTCAAGATGCATTTGCCTGCATTGGAAACC 9989

Qy       910 ---AlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLys 928
                ::::::         |||      |||    :::                |||:::|||
Db      9988 AAGAGTGTTGAACAGGAAGTAATTCAGTCACAACTAAGTCATTGTGTGAACTTGTATAAA 9929

Qy       929 GlyLeuSerArgLeuArg------------------------------------------ 934
                ||||||   ::::::
Db      9928 AGCCTGAGTGAAGTCAAGTCTGAAGTGGAAATGGTGATTAAAACCGGACGTCAAATTGTA 9869

Qy       935 ------------------------------------------------LeuProLeu 937
                                                             |||   |||
Db      9868 CAGAAAAAGCAGACAGAAAATCCCAAAGAGCTTGATGAACGAGTAACAGCTTTGAAATTG 9809

Qy       938 GluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHis 957
                |||       :::   |||                         |||   ||||||:::
Db      9808 CATTACAATGAGTTGGGTGCG---------------AAGGTAACAGAGAGAAAGCAACAG 9764

Qy       958 AlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAla 977
                   :::||||||          :::      :::|||:::|||
Db      9763 TTGGAGAAATGCTTGAAGTTGTCCCGTAAGATGAGAAAGGAA------------------ 9722

Qy       978 AlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHis 997
                                  :::::::::|||    |||:::
Db      9721 -----------------ATGAATGTCTTAACAGAATGG--------------------- 9701

Qy       998 LeuLeuAlaHisAspProAspTyrValLys-ValGlnAspIleGluGlnLeuLysAspVa 1017
             |||   |||   |||      :::         |||     |||      :::
Db      9700 CTGGCAGCAACAGATACAGAATTGACGAAGAGATCAGCAGTTG----------------- 9658

Qy      1017 lLysGluCys------LeuTrpPheValLeu---------GluIleLeuMetAlaLysAs 1032
              ||||||||        :::|||   :::              |||   |||:::   |||
Db      9657 -AAGGAATGCCAAGTAATTTGGATTCTGAAGTTGCCTGGGGAAAGGCTACTCAAAAAGAG 9599

Qy      1032 nGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAs 1052
                |||      |||    ::::::     |||       :::          |||
Db      9598 ATTGAGAAACAG-AAGGCTCACTTGAAGAGTGTTACAGAATTAGGAGAGTCTTTGAAAAT 9540

Qy      1052 pAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrVal----------- 1068
                |||     :::       :::     :::||||||       :::
Db      9539 GGTGTTGGGCAAGAAAGAAACCTTGGTAGAAGATAAACTGAGTCTTCTGAACAGTAACTG 9480

Qy      1069 ------------------CysAspValAlaMetAsnIleIleMetSer----------- 1078
                               :::     :::|||:::::::::
Db      9479 GATAGCTGTCACCTCCAGAGTAGAAGAATGGCTAAATCTTTTGTTGGAATACCAGAAACA 9420

Qy      1079 ----LysSerThrThrTyrSerLeuGluSerProLysAspProValLeuProAla---Ar 1096
                 ::::::            ::::::|||                 ::::::   |||
Db      9419 CATGGAAACCTTTGATCAGAACATAGAACAAATCACAAAGTGGATCATTCATGCAGATGA 9360

Qy      1096 gPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMe 1116
                   :::   :::|||                   |||      :::            ::
Db      9359 ACTTTTAGATGAGTCTGAAAAGAAGAAACCACAACAAAAGGAAGACATTCTTAAGCGTTT 9300

Qy      1116 tLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLy 1136
              :|||:::               :::||||||      :::        .             ||
Db      9299 AAAGGCTGAAATGAATGACATGCGCCCAAAGGTGGACTCCACACGTGACCAAGCAGCAAA 9240

Qy      1136 sProLeuSerSerAlaGly-----------------------LysGlnSerGlnThrLy 1148
              |    :::::::::     |||                              |||
Db      9239 ATTGATGGCAAACCGCGGTGACCACTGCAGGAAAGTAGTAGAGCCCCAAATCTCTGAGCT 9180

Qy      1148 sSerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGl 1168
                :::    |||         :::|||:::               :::            |||    |||
Db      9179 CAACCGTCGATTTGCAGCTATTTCTCACAGAATTAAGACTGGAAAGGCCTCCATTCCT-- 9122

Qy      1168 yArgIleLysGlyArgLeuAspSerSerGluMetAspHisSerGluAsnGluAspTyrTh 1188
                                        |||::::::               ::::::
Db      9121 ---------------------TTGAAGGAATTGGAGCAGTTTAACTCAGATATACAAAA 9084

Qy      1188 rMetSerSerProLeuProGlyLys----------------------------------- 1196
               :::      ||||||      :::
Db      9083 ATTGCTTGAACCACTGGAGGCTGAAATTCAGCAGGGGGTGAATCTGAAAGAGGAAGACTT 9024

Qy      1197 -LysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGl 1216
                         |||       :::|||::::::             :::||||||    :::
Db      9023 CAATAAAGATATGAGTGAAGACAATGAGGGTACTGTAAATGAATTGTTGCAAAGAGGAGA 8964

Qy      1216 yArgLysLysThrProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLy 1236
                      :::         :::|||::::::   ::::::            :::             ::
Db      8963 CAACTTACAACAAAGAATCACAGATGAGAGAAAGCGAGAGGAAATAAAGATAAAACAGCA 8904

Qy      1236 sLeuValGlnGluGln------------LysProLysGlySerGlnArgSerArgLysAr 1252
              :|||:::|||     :::                 |||          |||||||||    :::|||
Db      8903 GCTGTTACAGACAAAACATAATGCTCTCAAGGATTTGAGGTCTCAAAGAAGAAAAAAG-- 8846

Qy      1252 gGlyHisThrAlaSerGluSerAspGluGlnGlnTrpProGluGluLysArgLeuLysGl 1272
                          :::                         ||||||    :::    ||||||         ::
Db      8845 ------------GCCCTAGAAATTTCTCACCAGTGGTATCAGTACAAGAGGCAGGCTGA 8799

Qy      1272 uAspIleLeuGluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyAr 1292
              :|||:::|||:::          ||||||
Db      8798 TGATCTCCTGAAATGCTTGGATGAAATT-------------------------------- 8771

Qy      1292 gProProLysProLeuGlyGlyGlyThrProLysGluGluProThrMetLysThrSerLy 1312

Db      8771 ------------------------------------------------------------ 8771

Qy      1312 sLysGlySerLysLysLysSerGlyProProAlaProGluGluGlu---------GluGl 1329
                      ||||||    :::            |||      |||    :::|||                 :::||
Db      8770 -------GAAAAAAAATTAGCCAGCCTACCTGAACCCAGAGATGAAAGAAAATTAAAGGA 8718

Qy      1329 uGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgValSe 1349
                |      :::|||:::                  ||||||    |||    :::    :::    |||
Db      8717 AATTGATCGTGAATTG-----------CAGAAGAAGAAAGAGGAGCTGAATGCAGTGCG 8670

Qy      1349 rArgArgAlaGlnGlnArgAlaGluSerProGluSerSerAlaIleGluSerThrGln 1368
                |||:::|||:::         :::|||:::         :::    |||:::|||    ||||||
Db      8669 CAGGCAAGCTGAGGGCTTGTCTGAGAATGGGGCCGCAATGGCAGTGGAGCCAACTCAG 8612
```

RESULT 13
US-09-427-048A-10/c
; Sequence 10, Application US/09427048A
; Patent No. 6203975
; GENERAL INFORMATION:
; APPLICANT: Trustees of the University of Pennsylvania
; Wilson, James M.
; Fisher, Krishna J.
; Chen, Shu-Jen
; Weitzman, Matthew
; TITLE OF INVENTION: Improved Adenovirus Virus and
; Methods of Use Thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P O Box 457

```
;           CITY: Spring House
;           STATE: Pennsylvania
;           COUNTRY: USA
;           ZIP: 19477
;       COMPUTER READABLE FORM:
;           MEDIUM TYPE: Floppy disk
;           COMPUTER: IBM PC compatible
;           OPERATING SYSTEM: PC-DOS/MS-DOS
;           SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/09/427,048A
;           FILING DATE: 21-Oct-1999
;           CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: 08/836,022
;           FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;           NAME: Bak, Mary E.
;           REGISTRATION NUMBER: 31,215
;           REFERENCE/DOCKET NUMBER: GNVPN.008PCT
;       TELECOMMUNICATION INFORMATION:
;           TELEPHONE: 215-540-9200
;           TELEFAX: 215-540-5818
;   INFORMATION FOR SEQ ID NO: 10:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 19307 base pairs
;           TYPE: nucleic acid
;           STRANDEDNESS: double
;           TOPOLOGY: unknown
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-427-048A-10

Alignment Scores:
Pred. No.:              8.87e-06      Length:        19307
Score:                  184.50        Matches:       241
Percent Similarity:     33.38%        Conservative:  213
Best Local Similarity:  17.72%        Mismatches:    496
Query Match:            2.56%         Indels:        411
DB:                     3             Gaps:          50

US-09-512-581B-2 (1-1391) x US-09-427-048A-10 (1-19307)

Qy        206 AlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThr 225
              |||   :::  ||||||   :::         :::   |||    :::|||
Db      12049 GCTTCAGAACAACTGAACAGCCGGTGGACAGAATTCTGCCAATTGCTGAGTGAGAGAGTT 11990

Qy        226 Ala---GlnAlaIleGluProTyrIleThrThrPhePheAsnGlnValLeuMetLeuGly 244
                          :::      |||   ||||||:::||||||:::         |||
Db      11989 AACTGGCTAGAGTATCAAACCAACATCATTACCTTTTATAATCAGCTACAACAATTGGAA 11930

Qy        245 -----------------------------LysThrSerIleSerAspLeuSerGlu--- 253
                                           ||||||    :::  |||||||||
Db      11929 CAGATGACAACTACTGCCGAAAACTTGTTGAAAACCCAGTCTACCACCCTATCAGAGCCA 11870

Qy        254 HisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerVal 273
                                |||::::::          :::   :::   |||   ::::::
Db      11869 ACAGCAATTAAAAGCCAGTTAAAAATTTGTAAGGATGAAGTCAACAGATTGTCAGCTCTT 11810

Qy        274 LeuProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValVal 293
                 ||||||:::|||                                 :::  ::::::|||   :::
Db      11809 CAGCCTCAAATTGAG----------------------CAATTAAAAATTCAGAGTCTA 11774

Qy        294 LysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysPro 313
              :::|||      |||      |||                                  |||
Db      11773 CAACTGAAAGAAAAGGGACAGGGG-------------------------------CCA 11747

Qy        314 LeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCys 333
              ::::::                                     :::       :::
Db      11746 ATGTTT-------------------------------------CTGGATGCAGACTTT 11726

Qy        334 ValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGlu 353
              |||   |||   :::|||        ||||||
Db      11725 GTGGCCTTTACTAATCAT------TTTAACCAC------------------------- 11699

Qy        354 TyrLeuLysValArgSerHisAspProGluGluAlaIleArgHisAspValIleValSer 373
                                                                     |||
Db      11698 -------------------------------------------------ATCTTTGAT 11690

Qy        374 IleValThrAlaAlaLysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheVal 393
                 |||   |||   :::|||::::::        :::   |||   |||
Db      11689 GGTGTGAGGGCCAAAGAGAAAGAGCTACAGACAATTTTTGACACTTTACCACCAATGCGC 11630

Qy        394 ArgGluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMetMetGlyLeuAla 413
                 :::   |||:::                  :::|||
Db      11629 TATCAGGAGACAATGAGTAGC---------ATCAGG---------------------- 11603

Qy        414 GlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIle 433

Db      11603 ---------------------------------------------------------- 11603

Qy        434 AlaTrpIleLys------AspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAsp 451
                 ||||||:::          :::      |||   :::   |||   :::      :::
Db      11602 ACGTGGATCCAGCAGTCAGAAAGCAAACTCTCTGTACCTTATCTTAGTGTTACTGAATAT 11543

Qy        452 ArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrThr 471
                 ::::::   ||||||:::                :::      :::|||:::
Db      11542 GAAATAATGGAGGAGAGACTCGGGAAATTACAGGCTCTGCAAAGTTCTTTGAAA------ 11489

Qy        472 GluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAla 491
              |||:::            ||||||      |||:::            |||   ||||||
Db      11488 GAGCAACAAAATGGCTTCAACTATCTGAGTGACACTGTGAAGGAGATGGCCAAGAAAGCA 11429

Qy        492 LeuAsnGluMet------------------------------------------TrpLys 497
                 :::|||:::                                          ||||||
Db      11428 CCTTCAGAAATATGCCAGAAATATCTGTCAGAATTTGAAGAGATTGAGGGGCACTGGAAG 11369

Qy        498 ---------------------------CysGlnAsnLeuLeuArgHisGlnValLysAsp 508
                                         ||||||   |||         |||
Db      11368 AAACTTTCCTCCCAGTTGGTGGAAAGCTGCCAAAAGCTAGAAGAACAT----------- 11321

Qy        509 LeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLys 528
                    :::   |||      |||                :::|||   :::      |||
Db      11320 ---------ATGAATAAACTTCGAAAATTTCAGAATCACATAAAAACCTTA---CAGAAA 11273

Qy        529 ValMetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLys 548
                 |||                                :::         |||:::|||:::
Db      11272 TGGATG--------------------------GCTGAAGTTGATGTTTTCCTGAAAGAG 11240

Qy        549 PheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSer 568
                          |||   |||   |||   ::::::|||||||||:::
Db      11239 GAATGGCCTGCCCTGGGGGATGCTGAAATCCTGAAAAAACAGCTCAAA----------- 11192

Qy        569 ProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGly 588
                          |||:::         |||   ::::::   |||         |||
Db      11191 ---------CAATGCAGACTTTTAGTTGGTGATATTCAAACAATTCAGCCCAGTTTAAAT 11141

Qy        589 AsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIle 608
              :::                  |||         :::   ||||||         |||
Db      11140 AGT-----------GTTAATGAAGGTGGGCAGAAGATAAAGAGTGAAGCTGAACTTGAG 11093

Qy        609 AlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSer 628
                          ::::::||||||                  ::::::::::::|||
Db      11092 TTTGCATCCAGACTGGAGACAGAA---------------CTTAGAGAGCTTAACACTCAG 11048

Qy        629 IleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArgAla 648
                 |||            :::                  :::|||::::::|||
Db      11047 TGGGATCACATATGC---------CGCCAGGTCTACACCAGAAAGGAAGCCTTAAAGGCA 10997

Qy        649 GlyLeuGlu-----------LeuLeuLysValLeuSerPheThrHisProIleSerPhe 664
                 ||||||:::          |||   |||   ||||||         |||
Db      10996 GGTTTGGATAAAACCGTAAGCCTCCAAAAAGATCTATCAGAGATGCATGAGTGGATGACA 10937

Qy        665 HisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGlu------ 682
                 :::   |||   :::                           |||      ||||||
Db      10936 CAAGCTGAAGAAGAATATCTAGAGAGAGATTTGAATATAAAACTCCAGATGAATTACAG 10877
```

```
Qy      683 ---------------------LysValAlaGluAlaAlaLeuGln--------------- 690
                                 :::      |||   |||||||||
Db    10876 ACTGCTGTTGAAGAAATGAAGAGAGCTAAAGAAGAGGCACTACAAAAAGAAACTAAAGTG 10817

Qy      691 ------IlePheLysAsnThrGlySerLysIleGluGluAspPheProHisIleArgSer 708
                  :::   :::         |||   |||            |||      :::
Db    10816 AAACTCCTTACTGAGACTGTAAATAGTGTAATAGCTCACGCTCCACCCTCAGCACAAGAG 10757

Qy      709 AlaLeuLeuProValLeuHisHisLysSerLys--------------------------- 719
            ||||||         |||         :::
Db    10756 GCCTTAAAAAAGGAACTTGAAACTCTGACCACCAACTACCAATGGCTGTGCACCAGGCTG 10697

Qy      720 LysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSer 739
               |||      :::      :::            |||   |||   :::   |||
Db    10696 AATGGAAAATGCAAAACTTTGGAAGAAGTTTGGGCATGTTGGCATGAGTTATTGTCATAT 10637

Qy      740 LysGlu------------------------------------------------------ 741
               |||
Db    10636 TTAGAGAAAGCAAACAAGTGGCTCAATGAAGTAGAATTGAAACTTAAAACCATGGAAAAT 10577

Qy      742 ---------------ThrGlnPheAlaGlnIlePheGluProLeu--------------- 751
                              :::      ::::::   |||   |||
Db    10576 GTTCCTGCAGGACCTGAGGAAATCACTGAAGTGCTAGAATCTCTTGAAAATCTGATGCAT 10517

Qy      752 HisLysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGly 771
            |||         :::|||:::   :::      |||         |||         |||
Db    10516 CATTCAGAGGAGAACCCAAATCAGATTCGTCTATTGGCACAGACTCTTACAGATGGAGGA 10457

Qy      772 HisIleAlaLeuLeuAlaProAspGln---------PheAlaAlaProTrpLysSerTrp 788
               :::         |||      ::::::      |||   :::   |||:::
Db    10456 GTCATGGATGAACTGATCAATGAGGAGCTTGAGACGTTTAATTCTCGTTGGAGGGAACTA 10397

Qy      789 ValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThr 808
                        :::   |||      |||:::         :::         :::
Db    10396 CATGAAGAGGCTGTGAGGAAACAAAAGTTGCTTGAACAGAGTATCCAGTCTGCCCAGGAA 10337

Qy      809 ThrLys---------------------------LeuTrpValProAspGluGluValSer 819
               :::                           |||         |||::::::::::::
Db    10336 ATTGAAAAGTCCTTGCACTTAATTCAGGAGTCGCTTGAATTCATTGACAAGCAGTTGGCA 10277

Qy      820 ProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMet 839
                           |||:::   |||   :::|||
Db    10276 GCTTATATCACTGACAAGGTGGATGCAGCTCAAATGCCTCAGGAA------------GCC 10229

Qy      840 LysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHis 859
            :::         |||         :::   :::   |||                  |||
Db    10228 CAGAAAATCCAATCAGATTTGACAAGTCATGAGATAAGTTTAGAAGAAATGAAGAAACAT 10169

Qy      860 SerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArg 879
            :::   |||            |||      :::|||:::|||:::::::::      :::
Db    10168 AACCAGGGGAAGGATGCCAACCAAAGGGTTCTTTCACAAATTGATGTTGCACAGAAAAAA 10109

Qy      880 LeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGlu------ 897
            |||         |||         :::|||   |||:::|||         |||
Db    10108 TTACAAGATGTCTCCATGAAATTTCGATTATTCCAAAAACCAGCCAATTTTGAACAACGT 10049

Qy      898 ------------------IleIleThrLeuGluGlnTyrGlnLeuCysAlaLeu------ 909
                              |||:::         :::      |||   ||||||
Db    10048 CTAGAGGAAAGTAAGATGATTTTAGATGAAGTCAAGATGCATTTGCCTGCATTGGAAACC 9989

Qy      910 ---AlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLys 928
               ::::::      |||      |||   :::                  |||:::|||
Db     9988 AAGAGTGTTGAACAGGAAGTAATTCAGTCACAACTAAGTCATTGTGTGAACTTGTATAAA 9929

Qy      929 GlyLeuSerArgLeuArg------------------------------------------ 934
               ||||||   ::::::
Db     9928 AGCCTGAGTGAAGTCAAGTCTGAAGTGGAAATGGTGATTAAAACCGGACGTCAAATTGTA 9869

Qy      935 ---------------------------------------------------LeuProLeu 937
                                                               |||   |||
Db     9868 CAGAAAAAGCAGACAGAAAATCCCAAAGAGCTTGATGAACGAGTAACAGCTTTGAAATTG 9809

Qy      938 GluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHis 957
               |||   :::   |||                     |||   ||||||:::
Db     9808 CATTACAATGAGTTGGGTGCG---------------AAGGTAACAGAGAGAAAGCAACAG 9764

Qy      958 AlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAla 977
                     :::||||||   :::      :::|||:::|||
Db     9763 TTGGAGAAATGCTTGAAGTTGTCCCGTAAGATGAGAAAGGAA------------------ 9722

Qy      978 AlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHis 997
                              :::::::::|||   |||:::
Db     9721 ------------------ATGAATGTCTTAACAGAATGG--------------------- 9701

Qy      998 LeuLeuAlaHisAspProAspTyrValLys-ValGlnAspIleGluGlnLeuLysAspVa 1017
            |||   |||   |||   :::      |||    |||   :::
Db     9700 CTGGCAGCAACAGATACAGAATTGACGAAGAGATCAGCAGTTG----------------- 9658

Qy     1017 lLysGluCys------LeuTrpPheValLeu---------GluIleLeuMetAlaLysAs 1032
             |||||||||      :::|||   :::           |||   |||:::   |||
Db     9657 -AAGGAATGCCAAGTAATTTGGATTCTGAAGTTGCCTGGGGAAAGGCTACTCAAAAAGAG 9599

Qy     1032 nGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAs 1052
                   |||      |||   :::::: :::     |||      :::      |||
Db     9598 ATTGAGAAACAG-AAGGCTCACTTGAAGAGTGTTACAGAATTAGGAGAGTCTTTGAAAAT 9540

Qy     1052 pAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrVal----------- 1068
                   |||     :::       :::   :::||||||      :::
Db     9539 GGTGTTGGGCAAGAAAGAAACCTTGGTAGAAGATAAACTGAGTCTTCTGAACAGTAACTG 9480

Qy     1069 -------------------CysAspValAlaMetAsnIleIleMetSer----------- 1078
                                  :::      :::|||:::::::::
Db     9479 GATAGCTGTCACCTCCAGAGTAGAAGAATGGCTAAATCTTTTGTTGGAATACCAGAAACA 9420

Qy     1079 ----LysSerThrThrTyrSerLeuGluSerProLysAspProValLeuProAla---Ar 1096
                ::::::     :::::::|||                  ::::::   |||
Db     9419 CATGGAAACCTTTGATCAGAACATAGAACAAATCACAAAGTGGATCATTCATGCAGATGA 9360

Qy     1096 gPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMe 1116
                        :::   :::|||            |||      :::          ::
Db     9359 ACTTTTAGATGAGTCTGAAAAGAAGAAACCACAACAAAAGGAAGACATTCTTAAGCGTTT 9300

Qy     1116 tLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLy 1136
            :|||:::               :::||||||      :::                  ||
Db     9299 AAAGGCTGAAATGAATGACATGCGCCCAAAGGTGGACTCCACACGTGACCAAGCAGCAAA 9240

Qy     1136 sProLeuSerSerAlaGly------------------------LysGlnSerGlnThrLy 1148
             |  ::::::::    |||                           |||
Db     9239 ATTGATGGCAAACCGCGGTGACCACTGCAGGAAAGTAGTAGAGCCCCAAATCTCTGAGCT 9180

Qy     1148 sSerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGl 1168
             :::   |||            :::|||:::                  |||   |||
Db     9179 CAACCGTCGATTTGCAGCTATTTCTCACAGAATTAAGACTGGAAAGGCCTCCATTCCT-- 9122

Qy     1168 yArgIleLysGlyArgLeuAspSerSerGluMetAspHisSerGluAsnGluAspTyrTh 1188
                                           |||:::::       ::::::
Db     9121 ----------------------TTGAAGGAATTGGAGCAGTTTAACTCAGATATACAAAA 9084

Qy     1188 rMetSerSerProLeuProGlyLys----------------------------------- 1196
             :::      ||||||      :::
Db     9083 ATTGCTTGAACCACTGGAGGCTGAAATTCAGCAGGGGGTGAATCTGAAAGAGGAAGACTT 9024

Qy     1197 -LysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGl 1216
                   |||      :::|||:::::          :::||||||   :::
Db     9023 CAATAAAGATATGAGTGAAGACAATGAGGGTACTGTAAATGAATTGTTGCAAAGAGGAGA 8964

Qy     1216 yArgLysLysThrProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLy 1236
                   :::      :::|||:::::: ::::::        :::            ::
Db     8963 CAACTTACAACAAAGAATCACAGATGAGAGAAAGCGAGAGGAAATAAAGATAAAACAGCA 8904

Qy     1236 sLeuValGlnGluGln------------LysProLysGlySerGlnArgSerArgLysAr 1252
            :|||:::|||  :::            |||         ||||||||    :::|||
Db     8903 GCTGTTACAGACAAAACATAATGCTCTCAAGGATTTGAGGTCTCAAAGAAGAAAAAG-- 8846

Qy     1252 gGlyHisThrAlaSerGluSerAspGluGlnGlnTrpProGluGluLysArgLeuLysGl 1272
```

```
                         :::             ||||||   :::   ||||||      ::
Db       8845 -------------GCCCTAGAAATTTCTCACCAGTGGTATCAGTACAAGAGGCAGGCTGA 8799

Qy       1272 uAspIleLeuGluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyAr 1292
              :|||:::|||:::        ||||||
Db       8798 TGATCTCCTGAAATGCTTGGATGAAATT------------------------------- 8771

Qy       1292 gProProLysProLeuGlyGlyGlyThrProLysGluGluProThrMetLysThrSerLy 1312

Db       8771 ------------------------------------------------------------ 8771

Qy       1312 sLysGlySerLysLysLysSerGlyProProAlaProGluGluGlu---------GluGl 1329
                     ||||||   :::      |||   |||   :::|||          :::||
Db       8770 -------GAAAAAAAATTAGCCAGCCTACCTGAACCCAGAGATGAAAGAAAATTAAAGGA 8718

Qy       1329 uGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgValSe 1349
              |    :::|||:::                 ||||||   |||   :::   :::   |||
Db       8717 AATTGATCGTGAATTG-----------CAGAAGAAGAAAGAGGAGCTGAATGCAGTGCG 8670

Qy       1349 rArgArgAlaGlnGlnArgAlaGluSerProGluSerSerAlaIleGluSerThrGln 1368
                |||:::|||:::        :::|||:::        :::   |||:::|||   ||||||
Db       8669 CAGGCAAGCTGAGGGCTTGTCTGAGAATGGGGCCGCAATGGCAGTGGAGCCAACTCAG 8612
```

RESULT 14
US-08-195-487-3
; Sequence 3, Application US/08195487
; Patent No. 5783403
; GENERAL INFORMATION:
; APPLICANT: TOUKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,487
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/901,701
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6306 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..6306
; PUBLICATION INFORMATION:
; AUTHORS: COMPTON, DUANE A
; AUTHORS: SZILAK, ILLYA
; AUTHORS: CLEVELAND, DON W
; TITLE: PRIMARY STRUCTURE OF OF NUMA, AN INTRANUCLEAR
; TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
; TITLE: SEGREGATION OF PROTEINS AT MITOSIS
; JOURNAL: JOURNAL OF CELL BIOLOGY
; VOLUME: 116
; PAGES: 1395-1408
; DATE: MAR-1992
US-08-195-487-3

Alignment Scores:

Pred. No.:	1.9e-06	Length:	6306
Score:	182.50	Matches:	281
Percent Similarity:	31.34%	Conservative:	195
Best Local Similarity:	18.50%	Mismatches:	549
Query Match:	2.54%	Indels:	494
DB:	1	Gaps:	55

US-09-512-581B-2 (1-1391) x US-08-195-487-3 (1-6306)

```
Qy         16 ProProGlyValLysGluIleSerAspLysIleSerLysGluGluMetValArgArgLeu 35
              |||      ::::::|||:::   ||||||            |||   :::      ||||||
Db        823 CCCAAGGAGCTTGAGGAGCTGCGTGACAAG------AATGAGAGCCTTACCATGCGGCTG 876

Qy         36 LysMetValValLysThrPheMetAspMet------------------------------ 45
                    :::|||         |||:::
Db        877 CATGAAACCCTGAAGCAGTGCCAGGACCTGAAGACAGAGAAGAGCCAGATGGATCGCAAA 936

Qy         46 ---AspGlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeu---------AlaLeu 61
                 :::|||   |||||||||         :::|||         |||         |||
Db        937 ATCAACCAGCTTTCGGAGGAGAATGGAGACCTTTCCTTTAAGCTGCGGGAGTTTGCCAGT 996

Qy         62 HisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuValAla 81
              ||||||
Db        997 CATCTG------------------------------------------------------ 1002

Qy         82 CysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAsp 101

Db       1003 ---------------------------------------------------------CAG 1005

Qy        102 LysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLys 121
              :::|||:::|||                 :::|||                  |||   :::|||
Db       1006 CAGCTACAGGATGCCCTCAATGAGCTGACGGAG---------------GAGCACAGCAAG 1050

Qy        122 SerProGln------PheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLys 139
              :::   |||               :::         |||   :::   ::::::
Db       1051 GCCACTCAGGAGTGGCTAGAGAAGCAGGCCCAGCTGGAGAAGGAGCTCAGCGCAGCCCTG 1110

Qy        140 SerTyrAsnIleCysPheGluLeuGluAspSerAsnGluIle------------PheThr 155
                       |||   ||||||:::   |||||||||                    :::
Db       1111 CAGGACAAGAAATGC------CTTGAAGAGAAGAACGAAATCCTTCAGGGAAAACTTTCA 1164

Qy        156 GlnLeuTyrArgThrLeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMet 175
              ||||||         |||      :::   :::|||            :::|||      :::
Db       1165 CAGCTGGAAGAACACTTGTCCCAGCTGCAGGATAACCCACCCCAGGAGAAGGGCGAGGTG 1224

Qy        176 HisMetValAspLeuMetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeu 195
                                                    ||||||   :::      |||
Db       1225 ---------------------------------CTGGGTGATGTCTTGCAGCTGGAAACC 1251

Qy        196 LeuAspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyr 215
              |||               |||         :::      |||         :::
Db       1252 TTGAAGCAAGAGGCAGCCACTCTTGCTGCAAACAACACACAGCTCCAAGCCAGGGTAGAG 1311

Qy        216 AspLeu------------------------AlaLysAlaLeuLeuLysArg--------- 224
                 |||                         ||||||   |||   :::|||
Db       1312 ATGCTGGAGACTGAGCGAGGCCAGCAGGAAGCCAAGCTGCTTGCTGAGCGGGGCCACTTC 1371

Qy        225 ------ThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnValLeuMet 242
                          |||   :::         ||||||
Db       1372 GAAGAAGAAAAGCAGCAGCTGTCTAGCCTGATCACTGACCTG------------------ 1413

Qy        243 LeuGlyLysThrSerIleSerAspLeuSerGlu--------------------------- 253
                 ::::::|||||||||:::||||||:::
```

```
Db       1414 ------CAGAGCTCCATCTCCAACCTCAGCCAGGCCAAGGAAGAGCTGGAGCAGGCCTCC 1467

Qy        254 ------HisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeu------ 269
                    |||      |||       ::::::  ::::::  |||   |||
Db       1468 CAGGCTCATGGGGCCCGGTTGACTGCCCAGGTGGCCTCTCTGACCTCTGAGCTCACCACA 1527

Qy        270 LeuLeuSerValLeuProGlnLeuGluPheLys---------LeuLysSerAsnAspAsn 286
              |||   :::   :::   |||   :::   :::         ||||||
Db       1528 CTCAATGCCACCATCCAGCAACAGGATCAAGAACTGGCTGGCCTGAAGCAGCAGGCCAAA 1587

Qy        287 GluGluArgLeuGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGlu 306
              |||::::::   |||:::   :::   |||                 :::      |||
Db       1588 GAGAAGCAGGCCCAGCTAGCACAGACCCTC------------------CAACAGCAAGAA 1629

Qy        307 LeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHis 326
                 |||||||||
Db       1630 CAGGCCTCCCAGGGC--------------------------------------------- 1644

Qy        327 ValProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAsp 346
                    :::|||   :::      :::   :::|||
Db       1645 ------CTCCGCCACCAGGTGGAGCAGCTAAGCAGTAGCCTGAAGCAGAAGGAGCAGCAG 1698

Qy        347 LeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIle 366
              |||   |||::::::   |||                        :::||||||
Db       1699 TTG---AAGGAGGTAGCGGAG-----------------------AAGCAGGAGGCAACT 1731

Qy        367 ArgHisAspValIleValSerIleValThr------------------------------ 376
              |||   |||         :::   |||
Db       1732 AGGCAGGACCATGCCCAGCAACTGGCCACTGCTGCAGAGGAGCGAGAGGCCTCCTTAAGG 1791

Qy        377 ------------------------------------------------AlaAlaLysLys 380
                                                              |||||||||
Db       1792 GAGCGGGATGCGGCTCTCAAGCAGCTGGAGGCACTGGAGAAGGAGAAGGCTGCCAAGCTG 1851

Qy        381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
              :::||||||               :::   |||      |||:::
Db       1852 GAGATTCTGCAGCAGCAACTTCAGGTGGCTAATGAAGCCCGGGACAGTGCCCAGACCTCA 1911

Qy        401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyr--- 419
                    :::   ::::::|||               |||::::::   :::|||
Db       1912 GTGACACAGGCCCAGCGGGAGAAG------------GCAGAGCTGAGCCGGAAGGTGGAG 1959

Qy        420 AlaLeuGlnSerAlaAla-----------------GlyLysAspAlaAlaLysGlnIle 433
                 ||||||:::                        :::|||       |||:::
Db       1960 GAACTCCAGGCCTGTGTTGAGACAGCCCGCCAGGAACAGCATGAGGCCCAGGCCCAGGTT 2019

Qy        434 AlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeu 453
              |||   ::::::   :::|||               :::   :::      :::|||:::
Db       2020 GCAGAGCTAGAGTTGCAGCTGCGGTCTGAGCAGCAAAAAGCAACTGAGAAAGAAAGGGTG 2079

Qy        454 LeuValGluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArg 473
                    |||:::         |||      :::         |||:::   :::
Db       2080 GCCCAGGAGAAGGACCAGCTCCAGGAGCAGCTCCAGGCCCTCAAAGAGTCCTTGAAGGTC 2139

Qy        474 MetLysCysLeuTyrTyrLeuTyrAlaThrLeuAsp---------LeuAsnAlaValLys 490
                 |||                 :::|||:::               |||
Db       2140 ACCAAG------------------GGCAGCCTTGAAGAGGAGAAGCGCAGGGCTGCAGAT 2181

Qy        491 AlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeu 510
              ||||||   |||      :::|||   :::   |||:::   :::   :::   |||:::
Db       2182 GCCCTGGAAGAGCAGCAGCGTTGTATCTCTGAGCTGAAGGCAGAGACCCGAAGCCTGGTG 2241

Qy        511 AspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysValMet 530
              :::      |||:::   :::   :::
Db       2242 GAGCAGCATAAGCGGGAACGAAAGGAG--------------------------------- 2268

Qy        531 ValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPheThr 550

Db       2268 ------------------------------------------------------------ 2268

Qy        551 GlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerProThr 570
                    ||||||::::::         ||||||   ||||||         :::
Db       2269 ------CTGGAAGAAGAGAGGGCTGGGCGCAAGGGGCTGGAGGCTCGATTACTGCAGCTT 2322

Qy        571 CysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAsnPro 590
                       |||||||||         |||:::   :::
Db       2323 GGGGAGGCCCATCAGGCTGAG---------ACTGAAGTCCTGCGG--------------- 2358

Qy        591 LysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaPro 610
                                             :::   |||   |||   :::|||
Db       2359 -------------------------------CGGGAGCTGGCAGAGGCCATGGCTGCC 2385

Qy        611 ValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAsp 630
                 |||                     |||:::|||:::|||            |||
Db       2386 CAGCACACAGCTGAGAGTGAGTGTGAGCAGCTCGTCAAAGAAGTAGCTGCCTGGCGTGAC 2445

Qy        631 GlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeu 650
              |||      |||   :::   |||                     :::      |||
Db       2446 GGGTATGAGGATAGCCAGCAAGAGGAGGCACAGTATGGCGCCATGTTCCAGGAACAGCTG 2505

Qy        651 GluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAlaGluThrPhe 670
                 ||||||                                       |||
Db       2506 ATGACTTTGAAG-------------------------------------GAGGAATGT 2526

Qy        671 GluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAla-----------Glu 686
              |||         |||:::      ||||||||||||
Db       2527 GAGAAGGCCCGCCAGGAGCTGCAGGAGGCAAAGGAGAAGGTGGCAGGCATAGAATCCCAC 2586

Qy        687 AlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHisIle 706
              :::   |||||||||   :::         :::|||:::   |||         ::::::
Db       2587 AGCGAGCTCCAGATAAGCCGGCAGCAG---AACAAACTAGCTGAGCTCCATGCCAACCTG 2643

Qy        707 ArgSerAlaLeuLeuProValLeuHisHisLysSerLysLysGlyProProArgGlnAla 726
                 ||||||   |||
Db       2644 GCCAGAGCACTCCAGCAGGTC--------------------------------------- 2664

Qy        727 LysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGln 746
                                                   ||||||   :::      :::
Db       2665 -------------------------------CAAGAGAAGGAAGTCAGGGCCCAGAAG 2691

Qy        747 IlePheGluProLeuHis---------------------LysSerLeuAspProSer 758
              :::   :::   |||                         |||   :::   :::
Db       2692 CTTGCAGATGACCTCTCCACTCTGCAGGAAAAGATGGCTGCCACCAGCAAAGAGGTGGCC 2751

Qy        759 AsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaPro 778
                 ||||||   |||:::
Db       2752 CGCTTGGAGACCTTGGTC------------------------------------------ 2769

Qy        779 AspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAsp---Leu 797
              :::   |||         :::         :::      :::|||||||||:::
Db       2770 ---CGCAAGGCAGGTGAGCAGCAGGAAACAGCCTCCCGGGAGTTAGTCAAGGAGCCTGCG 2826

Qy        798 LeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGlu 817
                       ||||||   |||                  |||:::      ||||||
Db       2827 AGGGCAGGAGACAGACAGCCCGAG-----------------TGGCTG------GAAGAG 2862

Qy        818 ValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeu 837
                                   ||||||                           |||
Db       2863 CAACAGGGACGCCAGTTCTGCAGCACACAGGCAGCG-----------------CTGCAG 2904

Qy        838 GlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIle 857
                 |||:::               :::   |||         ||||||         :::
Db       2905 GCTATGGAGCGGGAGGCAGAGCAGATGGGCAATGAGCTGGAACGGCTGCGGGCCGCGCTG 2964

Qy        858 LeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArg 877
              :::   |||   |||         |||:::            :::   :::::::::|||
Db       2965 ATGGAGAGCCAGGGGCAGCAGCAGGAGGAGCGTGGGCAGCAGGAAAGGGAGGTGGCGCGG 3024

Qy        878 LeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGlu 897
              |||            |||   |||         ||||||   |||
Db       3025 CTGACCCAGGAGCGGGGCCGTGCCCAGGCTGACCTTGCCCTGGAG--------------- 3069

Qy        898 IleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGln 917

Db       3069 ------------------------------------------------------------ 3069
```

```
Qy        918 ValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeu 937

Db       3069 ------------------------------------------------------------ 3069

Qy        938 GluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHis 957
                                                        |||      ||||||
Db       3070 ------------------------------------------AAGGCGGCCAGAGCAGAG 3087

Qy        958 AlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAla 977
                          |||         :::|||   :::|||             :::  |||
Db       3088 CTTGAGATGCGGCTGCAGAACGCCCTCAACGAGCAGCGT---------GTGGAGTTCGCT 3138

Qy        978 AlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHis 997
                 :::   |||   |||         |||   |||
Db       3139 ACCCTGCAAGAGGCACTGGCTCATGCCCTGACGGAAAAGGAA------------------ 3180

Qy        998 LeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGlu------GlnLeuLys 1015
                          |||   :::      |||:::::    :::|||      |||:::|||
Db       3181 ------GGCAAGGACCAGGAGTTGGCCAAGCTTCGTGGTCTGGAGGCAGCCCAGATAAAA 3234

Qy       1016 AspValLysGluCysLeuTrpPheVal------LeuGluIleLeuMetAlaLysAsnGlu 1033
              ::::::::|||             |||                  :::||||||   |||
Db       3235 GAGCTGGAGGAACTTCGGCAAACCGTGAAGCAACTGAAGGAACAGCTGGCTAAGAAAGAA 3294

Qy       1034 AsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAla 1053
                          :::                  |||         :::|||:::
Db       3295 AAGGAGCACGCATCTGGCTCAGGAGCCCAATCTGAGGCTGCTGGCAGGACAGAG------ 3348

Qy       1054 GlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMet 1073
                    |||         |||:::                  |||      :::
Db       3349 ------CCAACAGGCCCCAAGCTGGAAGCACTGCGGGCAGAGGTGAGCAAGCTGGAACAG 3402

Qy       1074 AsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProValLeu 1093
                             :::            |||||||||         :::
Db       3403 CAATGCCAGAAGCAGCAGGAGCAGGCTGACAGCCTGGAACGCAGCCTCGAGGCTGAGCGG 3462

Qy       1094 ProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuPro 1113
                 :::|||         :::   |||                  :::      |||
Db       3463 GCCTCCCGG------GCTGAGCGGGACAGTGCTCTGGAGACTCTGCAGGGCCAGTTAGAG 3516

Qy       1114 ProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAla 1133
                 :::   :::                                       ||||||
Db       3517 GAGAAGGCCCAGGAG------------------------------------CTAGGGCAC 3540

Qy       1134 ValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThr-LysSerSerArgMetGl 1153
                          |||:::||||||   ::::::            :::   |||    ::
Db       3541 AGTCAGAGTGCCTTAGCCTCGGCCCAACGGGAGTTGGCTGCCTTCCGCACCAAGGTACAA 3600

Qy       1153 uThrValSerAsnAlaSerSerSerSerAsnProSerSerProGly---------ArgIl 1170
              :|||   :::            |||      |||      ||||||         |||::
Db       3601 GACCACAGCAAGGCTGAAGATGAGTGGAAGGCCCAGGTGGCCCGGGGCCGGCAAGAGGCT 3660

Qy       1170 eLysGlyArgLeuAspSerSer-GluMetAsp------------HisSerGluAsnGluA 1186
              :::::|||:::::   |||||| ::::::               |||   |||
Db       3661 GAGAGGAAAAATAGCCTCATCAGCAGCTTGGAGGAGGAGGTGTCCATCCTGAA------- 3713

Qy       1186 spTyrThrMetSerSerProLeuProGly-LysLysSerAspLysArgAspAspSerAsp 1205
                            ||||||   |||||| |||:::   :::    ::::::
Db       3714 --------------TCGCCAGGTCCTGGAGAAGGAGGGGGAGAGCAAGGAGTTGAAGCGG 3759

Qy       1206 LeuValArgSerGluLeuGluLysProArgGly------------------------Arg 1217
              ||||||   :::|||   ||||||   :::                           |||
Db       3760 CTGGTGATGGCCGAGTCAGAGAAGAGCCAGAAGCTGGAGGAGAGCTGCGCCTGCTGCAGG 3819

Qy       1218 LysLysThrProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeu 1237
              ::::::   |||   |||      |||   :::                        |||
Db       3820 CAGAGACAGCCAGCAACAGTGCCAGAGCTGCAG---------------AACGCAGCTCTG 3864

Qy       1238 ValGlnGluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSer 1257
              :::      ::::::   :::   |||   |||      ::::::         ||||||
Db       3865 CTCTGCGGGAGGAGGTGCAGAGCCTCCGGGAGGGAGGCTGAGAAACAGCGGGTGGCTTCA 3924

Qy       1258 GluSerAspGluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsn 1277
              |||:::      |||:::         :::            |||::::::
Db       3925 GAGAACCTGCGGCAGGAGCTGACCTCACAGGCTGAGCGTGCGGAGGAGCTG------GGC 3978

Qy       1278 GluAspGluGlnAsnSerProProLys-LysGlyLysArgGlyArgProProLysProLe 1297
              ::::::   :::            |||       ::::::   ||||||     |||
Db       3979 CAAGAATTGAAGGCGTGGCAGGAGAAGTTCTTCCAGAAAGAGCAGGCCCTCTCCACCCTG 4038

Qy       1297 uGlyGlyGlyThrProLysGluGluPro------------ThrMetLysThrSerLysLy 1313
                        ||||||         |||                  :::   |||
Db       4039 CAGCTCGAGCACACCAGCACACAGGCCCTGGTGAGTGAGCTGCTGCCAGCTAAGCACCTC 4098

Qy       1313 sGlySerLysLysLysSerGlyProProAlaProGluGluGluGluGluGlu-------- 1330
              |||    :::               |||   |||
Db       4099 TGCCAGCAGCTGCAGGCCGAGCAGGCCGCTGCCGAGAAACGCCACCGTGAGGAGCTGGAG 4158

Qy       1330 ------------------------------------------------------------ 1330

Db       4159 CAGAGCAAGCAGGCCGCTGGGGGACTGCGGGCAGAGCTGCTGCGGGCCCAGCGGGAGCTT 4218

Qy       1331 ---------------------GluArgGlnSerGlyAsnThrGluGlnLysSerLysSe 1343
                                       |||            |||||||||   |||
Db       4219 GGGGAGCTGATTCCTCTGCGGCAGAAGGTGGCAGAGCAGGACCGAACAGCTCAGCAGCTG 4278

Qy       1343 rLysGlnHisArgValSerArg-ArgAlaGlnGlnArgAlaGluSerProGluSerSerA 1363
                  ||||  |||   :::    ||||||   :::   ||||||               |
Db       4279 CGGGCAGAGAAGGCCAGCTATGCAGAGCAGCTGAGCATGCTGAAGAAGGCGCATGGCCTG 4338

Qy       1363 laIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgPro 1378
              ||                              |||||||||   |||
Db       4339 CT------------------------------GGCAGAGGAGAACCG 4355
```

RESULT 15
PCT-US93-06160-3
; Sequence 3, Application PC/TUS9306160
; GENERAL INFORMATION:
;   APPLICANT:
;   TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;   TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: TESTA HURWITZ & THIBEAULT
;   STREET: 53 STATE STREET
;   CITY: BOSTON
;   STATE: MA
;   COUNTRY: USA
;   ZIP: 02109
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US93/06160
;   FILING DATE: 19930621
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: PITCHER ESQ, EDMUND R
;   REGISTRATION NUMBER: 27,829
;   REFERENCE/DOCKET NUMBER: MTP-013
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 617/248-7000
;   TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 6306 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..6306
;   PUBLICATION INFORMATION:

```
;       AUTHORS:  COMPTON, DUANE A
;       AUTHORS:  SZILAK, ILLYA
;       AUTHORS:  CLEVELAND, DON W
;       TITLE:  PRIMARY STRUCTURE OF OF NUMA, AN INTRANUCLEAR
;       TITLE:  PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
;       TITLE:  SEGREGATION OF PROTEINS AT MITOSIS
;       JOURNAL:  JOURNAL OF CELL BIOLOGY
;       VOLUME:  116
;       PAGES:  1395-1408
;       DATE:  MAR-1992
PCT-US93-06160-3

Alignment Scores:
Pred. No.:               1.9e-06        Length:         6306
Score:                   182.50         Matches:        281
Percent Similarity:      31.34%         Conservative:   195
Best Local Similarity:   18.50%         Mismatches:     549
Query Match:             2.54%          Indels:         494
DB:                      5              Gaps:           55

US-09-512-581B-2 (1-1391) x PCT-US93-06160-3 (1-6306)

Qy          16 ProProGlyValLysGluIleSerAspLysIleSerLysGluGluMetValArgArgLeu 35
               |||       ::::::|||:::    ||||||          |||     :::        ||||||
Db         823 CCCAAGGAGCTTGAGGAGCTGCGTGACAAG------AATGAGAGCCTTACCATGCGGCTG 876

Qy          36 LysMetValValLysThrPheMetAspMet----------------------------- 45
                         :::|||          |||:::
Db         877 CATGAAACCCTGAAGCAGTGCCAGGACCTGAAGACAGAGAAGAGCCAGATGGATCGCAAA 936

Qy          46 ---AspGlnAspSerGluGluLysGluLeuTyrLeuAsnLeu---------AlaLeu 61
                  :::|||   |||||||||       :::|||          |||          |||
Db         937 ATCAACCAGCTTTCGGAGGAGAATGGAGACCTTTCCTTTAAGCTGCGGGAGTTTGCCAGT 996

Qy          62 HisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuValAla 81
               ||||||
Db         997 CATCTG------------------------------------------------------ 1002

Qy          82 CysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerProAsp 101

Db        1003 ---------------------------------------------------------CAG 1005

Qy         102 LysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLys 121
               :::|||:::|||                :::|||                |||    :::|||
Db        1006 CAGCTACAGGATGCCCTCAATGAGCTGACGGAG---------------GAGCACAGCAAG 1050

Qy         122 SerProGln------PheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLys 139
               :::    |||            :::          |||    :::     ::::::
Db        1051 GCCACTCAGGAGTGGCTAGAGAAGCAGGCCCAGCTGGAGAAGGAGCTCAGCGCAGCCCTG 1110

Qy         140 SerTyrAsnIleCysPheGluLeuGluAspSerAsnGluIle-----------PheThr 155
                           |||      ||||||:::   |||||||||                  :::
Db        1111 CAGGACAAGAAATGC------CTTGAAGAGAAGAACGAAATCCTTCAGGGAAAACTTTCA 1164

Qy         156 GlnLeuTyrArgThrLeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMet 175
               ||||||          |||          :::      :::|||            :::|||      :::
Db        1165 CAGCTGGAAGAACACTTGTCCCAGCTGCAGGATAACCCACCCCAGGAGAAGGGCGAGGTG 1224

Qy         176 HisMetValAspLeuMetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeu 195
                                                     ||||||     :::          |||
Db        1225 ---------------------------------CTGGGTGATGTCTTGCAGCTGGAAACC 1251

Qy         196 LeuAspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyr 215
               |||                |||            :::      |||        :::
Db        1252 TTGAAGCAAGAGGCAGCCACTCTTGCTGCAAACAACACACAGCTCCAAGCCAGGGTAGAG 1311

Qy         216 AspLeu-----------------------AlaLysAlaLeuLeuLysArg--------- 224
                  |||                           ||||||   |||    :::|||
Db        1312 ATGCTGGAGACTGAGCGAGGCCAGCAGGAAGCCAAGCTGCTTGCTGAGCGGGGCCACTTC 1371

Qy         225 ------ThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnValLeuMet 242
                             |||    :::             ||||||
Db        1372 GAAGAAGAAAAGCAGCAGCTGTCTAGCCTGATCACTGACCTG------------------ 1413

Qy         243 LeuGlyLysThrSerIleSerAspLeuSerGlu-------------------------- 253
                     :::::|||||||||:::||||||:::
Db        1414 ------CAGAGCTCCATCTCCAACCTCAGCCAGGCCAAGGAAGAGCTGGAGCAGGCCTCC 1467

Qy         254 ------HisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeu------ 269
                        |||        |||        ::::::   ::::::    |||    |||
Db        1468 CAGGCTCATGGGGCCCGGTTGACTGCCCAGGTGGCCTCTCTGACCTCTGAGCTCACCACA 1527

Qy         270 LeuLeuSerValLeuProGlnLeuGluPheLys---------LeuLysSerAsnAspAsn 286
               |||    :::    :::    |||    :::    :::               ||||||
Db        1528 CTCAATGCCACCATCCAGCAACAGGATCAAGAACTGGCTGGCCTGAAGCAGCAGGCCAAA 1587

Qy         287 GluGluArgLeuGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGlu 306
               |||:::::    |||:::     :::    |||                          :::        |||
Db        1588 GAGAAGCAGGCCCAGCTAGCACAGACCCTC-----------------CAACAGCAAGAA 1629

Qy         307 LeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHis 326
                  |||||||||
Db        1630 CAGGCCTCCCAGGGC--------------------------------------------- 1644

Qy         327 ValProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAsp 346
                     :::|||      :::          :::      :::|||
Db        1645 ------CTCCGCCACCAGGTGGAGCAGCTAAGCAGTAGCCTGAAGCAGAAGGAGCAGCAG 1698

Qy         347 LeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIle 366
               |||    |||::::::     |||                                        :::||||||
Db        1699 TTG---AAGGAGGTAGCGGAG-----------------------AAGCAGGAGGCAACT 1731

Qy         367 ArgHisAspValIleValSerIleValThr----------------------------- 376
               |||    |||                   :::    |||
Db        1732 AGGCAGGACCATGCCCAGCAACTGGCCACTGCTGCAGAGGAGCGAGAGGCCTCCTTAAGG 1791

Qy         377 --------------------------------------------AlaAlaLysLys 380
                                                               |||||||||
Db        1792 GAGCGGGATGCGGCTCTCAAGCAGCTGGAGGCACTGGAGAAGGAGAAGGCTGCCAAGCTG 1851

Qy         381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
               :::||||||                  :::    |||      |||:::
Db        1852 GAGATTCTGCAGCAGCAACTTCAGGTGGCTAATGAAGCCCGGGACAGTGCCCAGACCTCA 1911

Qy         401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyr--- 419
                     :::    ::::::|||                 |||::::::    :::|||
Db        1912 GTGACACAGGCCCAGCGGGAGAAG-----------GCAGAGCTGAGCCGGAAGGTGGAG 1959

Qy         420 AlaLeuGlnSerAlaAla-----------------GlyLysAspAlaAlaLysGlnIle 433
                  ||||||:::                                :::|||       |||:::
Db        1960 GAACTCCAGGCCTGTGTTGAGACAGCCCGCCAGGAACAGCATGAGGCCCAGGCCCAGGTT 2019

Qy         434 AlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeu 453
                  |||    ::::::       :::|||                :::    :::     :::|||:::
Db        2020 GCAGAGCTAGAGTTGCAGCTGCGGTCTGAGCAGCAAAAAGCAACTGAGAAAGAAAGGGTG 2079

Qy         454 LeuValGluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArg 473
                     |||:::           |||        :::                |||:::     :::
Db        2080 GCCCAGGAGAAGGACCAGCTCCAGGAGCAGCTCCAGGCCCTCAAAGAGTCCTTGAAGGTC 2139

Qy         474 MetLysCysLeuTyrTyrLeuTyrAlaThrLeuAsp---------LeuAsnAlaValLys 490
                  |||                      :::|||:::                  |||
Db        2140 ACCAAG-----------------GGCAGCCTTGAAGAGGAGAAGCGCAGGGCTGCAGAT 2181

Qy         491 AlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeu 510
               ||||||    |||          :::|||     :::       |||:::     :::     :::    |||:::
Db        2182 GCCCTGGAAGAGCAGCAGCGTTGTATCTCTGAGCTGAAGGCAGAGACCCGAAGCCTGGTG 2241

Qy         511 AspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysValMet 530
               :::          |||:::     :::     :::
Db        2242 GAGCAGCATAAGCGGGAACGAAAGGAG--------------------------------- 2268

Qy         531 ValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPheThr 550

Db        2268 ------------------------------------------------------------ 2268
```

```
Qy    551 GlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerProThr 570
Db   2269 ------CTGGAAGAAGAGAGGGCTGGGCGCAAGGGGCTGGAGGCTCGATTACTGCAGCTT 2322
Qy    571 CysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAsnPro 590
Db   2323 GGGGAGGCCCATCAGGCTGAG---------ACTGAAGTCCTGCGG--------------- 2358
Qy    591 LysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaPro 610
Db   2359 --------------------------------CGGGAGCTGGCAGAGGCCATGGCTGCC 2385
Qy    611 ValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAsp 630
Db   2386 CAGCACACAGCTGAGAGTGAGTGTGAGCAGCTCGTCAAAGAAGTAGCTGCCTGGCGTGAC 2445
Qy    631 GlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeu 650
Db   2446 GGGTATGAGGATAGCCAGCAAGAGGAGGCACAGTATGGCGCCATGTTCCAGGAACAGCTG 2505
Qy    651 GluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAlaGluThrPhe 670
Db   2506 ATGACTTTGAAG--------------------------------------GAGGAATGT 2526
Qy    671 GluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAla-----------Glu 686
Db   2527 GAGAAGGCCCGCCAGGAGCTGCAGGAGGCAAAGGAGAAGGTGGCAGGCATAGAATCCCAC 2586
Qy    687 AlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHisIle 706
Db   2587 AGCGAGCTCCAGATAAGCCGGCAGCAG---AACAAACTAGCTGAGCTCCATGCCAACCTG 2643
Qy    707 ArgSerAlaLeuLeuProValLeuHisHisLysSerLysLysGlyProProArgGlnAla 726
Db   2644 GCCAGAGCACTCCAGCAGGTC---------------------------------------- 2664
Qy    727 LysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGln 746
Db   2665 --------------------------------CAAGAGAAGGAAGTCAGGGCCCAGAAG 2691
Qy    747 IlePheGluProLeuHis-----------------------LysSerLeuAspProSer 758
Db   2692 CTTGCAGATGACCTCTCCACTCTGCAGGAAAAGATGGCTGCCACCAGCAAAGAGGTGGCC 2751
Qy    759 AsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaPro 778
Db   2752 CGCTTGGAGACCTTGGTC------------------------------------------ 2769
Qy    779 AspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAsp---Leu 797
Db   2770 ---CGCAAGGCAGGTGAGCAGCAGGAAACAGCCTCCCGGGAGTTAGTCAAGGAGCCTGCG 2826
Qy    798 LeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGlu 817
Db   2827 AGGGCAGGAGACAGACAGCCCGAG-----------------TGGCTG------GAAGAG 2862
Qy    818 ValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeu 837
Db   2863 CAACAGGGACGCCAGTTCTGCAGCACACAGGCAGCG-----------------CTGCAG 2904
Qy    838 GlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIle 857
Db   2905 GCTATGGAGCGGGAGGCAGAGCAGATGGGCAATGAGCTGGAACGGCTGCGGGCCGCGCTG 2964
Qy    858 LeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArg 877
Db   2965 ATGGAGAGCCAGGGGCAGCAGCAGGAGGAGCGTGGGCAGCAGGAAAGGGAGGTGGCGCGG 3024
Qy    878 LeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGlu 897
Db   3025 CTGACCCAGGAGCGGGGCCGTGCCCAGGCTGACCTTGCCCTGGAG--------------- 3069
Qy    898 IleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGln 917
Db   3069 ------------------------------------------------------------ 3069
Qy    918 ValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeu 937
Db   3069 ------------------------------------------------------------ 3069
Qy    938 GluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHis 957
Db   3070 -----------------------------------------AAGGCGGCCAGAGCAGAG 3087
Qy    958 AlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAla 977
Db   3088 CTTGAGATGCGGCTGCAGAACGCCCTCAACGAGCAGCGT---------GTGGAGTTCGCT 3138
Qy    978 AlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHis 997
Db   3139 ACCCTGCAAGAGGCACTGGCTCATGCCCTGACGGAAAAGGAA------------------ 3180
Qy    998 LeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGlu------GlnLeuLys 1015
Db   3181 ------GGCAAGGACCAGGAGTTGGCCAAGCTTCGTGGTCTGGAGGCAGCCCAGATAAAA 3234
Qy   1016 AspValLysGluCysLeuTrpPheVal------LeuGluIleLeuMetAlaLysAsnGlu 1033
Db   3235 GAGCTGGAGGAACTTCGGCAAACCGTGAAGCAACTGAAGGAACAGCTGGCTAAGAAAGAA 3294
Qy   1034 AsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAla 1053
Db   3295 AAGGAGCACGCATCTGGCTCAGGAGCCCAATCTGAGGCTGCTGGCAGGACAGAG------ 3348
Qy   1054 GlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMet 1073
Db   3349 ------CCAACAGGCCCCAAGCTGGAAGCACTGCGGGCAGAGGTGAGCAAGCTGGAACAG 3402
Qy   1074 AsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProValLeu 1093
Db   3403 CAATGCCAGAAGCAGCAGGAGCAGGCTGACAGCCTGGAACGCAGCCTCGAGGCTGAGCGG 3462
Qy   1094 ProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuPro 1113
Db   3463 GCCTCCCGG------GCTGAGCGGGACAGTGCTCTGGAGACTCTGCAGGGCCAGTTAGAG 3516
Qy   1114 ProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAla 1133
Db   3517 GAGAAGGCCCAGGAG-----------------------------------CTAGGGCAC 3540
Qy   1134 ValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThr-LysSerSerArgMetGl 1153
Db   3541 AGTCAGAGTGCCTTAGCCTCGGCCCAACGGGAGTTGGCTGCCTTCCGCACCAAGGTACAA 3600
Qy   1153 uThrValSerAsnAlaSerSerSerSerAsnProSerSerProGly---------ArgIl 1170
Db   3601 GACCACAGCAAGGCTGAAGATGAGTGGAAGGCCCAGGTGGCCCGGGGCCGGCAAGAGGCT 3660
Qy   1170 eLysGlyArgLeuAspSerSer-GluMetAsp-----------HisSerGluAsnGluA 1186
Db   3661 GAGAGGAAAAATAGCCTCATCAGCAGCTTGGAGGAGGAGGTGTCCATCCTGAA------- 3713
Qy   1186 spTyrThrMetSerSerProLeuProGly-LysLysSerAspLysArgAspAspSerAsp 1205
Db   3714 -------------TCGCCAGGTCCTGGAGAAGGAGGGGGAGAGCAAGGAGTTGAAGCGG 3759
Qy   1206 LeuValArgSerGluLeuGluLysProArgGly-----------------------Arg 1217
Db   3760 CTGGTGATGGCCGAGTCAGAGAAGAGCCAGAAGCTGGAGGAGAGCTGCGCCTGCTGCAGG 3819
Qy   1218 LysLysThrProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeu 1237
Db   3820 CAGAGACAGCCAGCAACAGTGCCAGAGCTGCAG--------------AACGCAGCTCTG 3864
Qy   1238 ValGlnGluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSer 1257
```

```
Db      3865 CTCTGCGGGAGGAGGTGCAGAGCCTCCGGGAGGGAGGCTGAGAAACAGCGGGTGGCTTCA 3924

Qy      1258 GluSerAspGluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsn 1277
             |||:::      |||:::         :::            |||::::::
Db      3925 GAGAACCTGCGGCAGGAGCTGACCTCACAGGCTGAGCGTGCGGAGGAGCTG------GGC 3978

Qy      1278 GluAspGluGlnAsnSerProProLys-LysGlyLysArgGlyArgProProLysProLe 1297
             ::::::   :::            |||         ::::::   ||||||      |||
Db      3979 CAAGAATTGAAGGCGTGGCAGGAGAAGTTCTTCCAGAAAGAGCAGGCCCTCTCCACCCTG 4038

Qy      1297 uGlyGlyGlyThrProLysGluGluPro------------ThrMetLysThrSerLysLy 1313
                       ||||||         |||                  :::   |||
Db      4039 CAGCTCGAGCACACCAGCACACAGGCCCTGGTGAGTGAGCTGCTGCCAGCTAAGCACCTC 4098

Qy      1313 sGlySerLysLysLysSerGlyProProAlaProGluGluGluGluGluGlu-------- 1330
                 |||      :::           |||   |||
Db      4099 TGCCAGCAGCTGCAGGCCGAGCAGGCCGCTGCCGAGAAACGCCACCGTGAGGAGCTGGAG 4158

Qy      1330 ------------------------------------------------------------ 1330

Db      4159 CAGAGCAAGCAGGCCGCTGGGGACTGCGGGCAGAGCTGCTGCGGGCCCAGCGGGAGCTT 4218

Qy      1331 ---------------------GluArgGlnSerGlyAsnThrGluGlnLysSerLysSe 1343
                                      |||            |||||||||   |||
Db      4219 GGGGAGCTGATTCCTCTGCGGCAGAAGGTGGCAGAGCAGGACCGAACAGCTCAGCAGCTG 4278

Qy      1343 rLysGlnHisArgValSerArg-ArgAlaGlnGlnArgAlaGluSerProGluSerSerA 1363
                 |||   |||   :::    ||||||   :::   ||||||                     |
Db      4279 CGGGCAGAGAAGGCCAGCTATGCAGAGCAGCTGAGCATGCTGAAGAAGGCGCATGGCCTG 4338

Qy      1363 laIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgPro 1378
             ||                              |||||||||   |||
Db      4339 CT-----------------------------GGCAGAGGAGAACCG 4355
```

Search completed: September 25, 2003, 01:48:10
Job time : 512 secs

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:          September 24, 2003, 20:21:35 ; Search time 769 Seconds
                                                (without alignments)
                                                4882.859 Million cell updates/sec

Title:           US-09-512-581B-2
Perfect score:   7193
Sequence:        1 MAHSKTRTNDGKITYPPGVK..........QKGRGRPSKTPSPSQPKKNV 1391

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters:      5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09512581/runat_23092003_163546_5464/app_query.fasta_1.1543
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09512581_@CGN_1_1_0_@runat_23092003_163546_5464 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :       N_Geneseq_19Jun03:*
                 1:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
                 2:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
                 3:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
                 4:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
                 5:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
                 6:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
                 7:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
                 8:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
                 9:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
                 10:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
                 11:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
                 12:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
                 13:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
                 14:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
                 15:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
                 16:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
                 17:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
                 18:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
                 19:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
                 20:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
                 21:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
                 22:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
                 23:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
                 24:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
                 25:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	7193	100.0	5271	21	AAA28051	Human androgen shu
	2	7193	100.0	5355	21	AAA28052	Human androgen shu
	3	7155	99.5	7473	23	ABV22430	Human prostate exp
	4	7149	99.4	7473	23	ABV28244	Human prostate exp
	5	7148	99.4	7473	23	ABV25469	Human prostate exp
	6	4760.5	66.2	6744	25	ABX14056	cDNA encoding huma
	7	3047	42.4	5469	24	ABQ99271	Human coding seque
	8	2932.5	40.8	3957	21	AAA47423	Sequence encoding
	9	1986.5	27.6	3920	23	ABL22105	Drosophila melanog
c	10	1729.5	24.0	6990	23	ABL22104	Drosophila melanog
	11	1409.5	19.6	2496	24	ABK35358	Human cDNA encodin
	12	967	13.4	1104	24	ABS56684	Euchromosome fragi
	13	912.5	12.7	7943	22	AAK85948	Human immune/haema
	14	864	12.0	714	22	AAS44727	Human full-length
	15	762.5	10.6	1217	21	AAF18338	Lung cancer associ
	16	748.5	10.4	2295	22	AAS02396	Human secreted pro
	17	706.5	9.8	2212	22	AAH17132	Human cDNA sequenc
	18	698	9.7	439	20	AAV87629	EST clone DY17. H
	19	645	9.0	772	22	AAH03894	Human cDNA clone (
	20	609	8.5	738	20	AAZ15259	Human gene express
	21	591	8.2	461	24	ABL89669	Human polynucleoti
	22	513	7.1	351	21	AAZ42861	Human 5' EST isola
	23	501	7.0	295	21	AAC02817	Human secreted pro
	24	375.5	5.2	1489	22	AAH18066	Human cDNA sequenc
	25	370	5.1	785	22	AAH07766	Human cDNA clone (
	26	357	5.0	227	24	ABQ94421	Tumour suppression
c	27	349	4.9	530	21	AAZ80598	Human colon cancer
	28	276	3.8	433	22	AAS44899	Human contig polyn
	29	248.5	3.5	9361	23	ABV25498	Human prostate exp
	30	235	3.3	20448	23	ABL19989	Drosophila melanog
	31	232	3.2	8493	25	ACA03930	cDNA downregulated
c	32	232	3.2	8527	22	AAI60039	Human polynucleoti
	33	231.5	3.2	9274	24	ABK13224	DNA encoding novel
	34	231	3.2	8503	22	AAI58253	Human polynucleoti
	35	230	3.2	12777	23	ABL28641	Drosophila melanog
	36	229	3.2	7992	24	ABK32895	DNA encoding human
	37	229	3.2	8083	23	AAS70692	DNA encoding novel
	38	228	3.2	7034	22	AAH19496	Murine coding sequ
	39	225	3.1	5334	19	AAX14598	H. pylori GHPO 175
c	40	223	3.1	557	24	ABK53631	Human eosinophil-m
	41	222.5	3.1	11116	22	AAI59270	Human polynucleoti
	42	220.5	3.1	7713	22	AAK51958	Human polynucleoti
	43	216.5	3.0	4534	23	ABV22145	Human prostate exp
	44	216.5	3.0	4534	23	ABV27985	Human prostate exp
	45	216.5	3.0	19303	23	ABL22098	Drosophila melanog

ALIGNMENTS

```
RESULT 1
AAA28051
ID   AAA28051 standard; cDNA; 5271 BP.
XX
AC   AAA28051;
XX
DT   01-DEC-2000  (first entry)
XX
DE   Human androgen shutoff gene 3 (AS3) cDNA sequence.
XX
KW   Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;
KW   chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;
KW   diagnosis; treatment; cytostatic; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
```

```
FT   exon            1..46
FT                   /*tag=  a
FT                   /number= 1
FT   exon            47..173
FT                   /*tag=  b
FT                   /number= 2
FT   CDS             66..4241
FT                   /*tag=  c
FT                   /product= "AS3"
FT                   /note= "Androgen shutoff gene 3 protein, the CDS is
FT                           specifically claimed as SEQ ID #3"
FT   exon            174..377
FT                   /*tag=  d
FT                   /number= 3
FT   exon            378..464
FT                   /*tag=  e
FT                   /number= 4
FT   exon            465..562
FT                   /*tag=  f
FT                   /number= 5
FT   exon            563..689
FT                   /*tag=  e
FT                   /number= 6
FT   exon            690..854
FT                   /*tag=  f
FT                   /number= 7
FT   exon            855..911
FT                   /*tag=  g
FT                   /number= 8
FT   exon            912..1027
FT                   /*tag=  h
FT                   /number= 9
FT   exon            1028..1122
FT                   /*tag=  i
FT                   /number= 10
FT   exon            1123..1268
FT                   /*tag=  j
FT                   /number= 11
FT   exon            1269..1420
FT                   /*tag=  k
FT                   /number= 12
FT   exon            1421..1534
FT                   /*tag=  l
FT                   /number= 13
FT   exon            1535..1616
FT                   /*tag=  m
FT                   /number= 14
FT   exon            1617..1665
FT                   /*tag=  n
FT                   /number= 15
FT   exon            1666..1805
FT                   /*tag=  o
FT                   /number= 16
FT   exon            1806..1921
FT                   /*tag=  p
FT                   /number= 17
FT   exon            1922..2027
FT                   /*tag=  q
FT                   /number= 18
FT   exon            2028..2188
FT                   /*tag=  r
FT                   /number= 19
FT   exon            2189..2312
FT                   /*tag=  s
FT                   /number= 20
FT   exon            2313..2471
FT                   /*tag=  t
FT                   /number= 21
FT   exon            2472..2540
FT                   /*tag=  u
FT                   /number= 22
FT   exon            2541..2677
FT                   /*tag=  v
FT                   /number= 23
FT   exon            2678..2801
FT                   /*tag=  w
FT                   /number= 24
FT   exon            2802..3006
FT                   /*tag=  x
FT                   /number= 25
FT   exon            3007..3121
FT                   /*tag=  y
FT                   /number= 26
FT   exon            3122..3254
FT                   /*tag=  z
FT                   /number= 27
FT   exon            3255..3374
FT                   /*tag=  aa
FT                   /number= 28
FT   exon            3375..3437
FT                   /*tag=  ab
FT                   /number= 29
FT   exon            3438..3583
FT                   /*tag=  ac
FT                   /number= 30
FT   exon            3584..3689
FT                   /*tag=  ad
FT                   /number= 31
FT   exon            3690..4129
FT                   /*tag=  ae
FT                   /number= 32
FT   exon            4130..4354
FT                   /*tag=  af
FT                   /number= 33
FT   misc_signal     4320..4324
FT                   /*tag=  ag
FT                   /note= "Destabilizing signal"
FT   exon            4355..5253
FT                   /*tag=  ah
FT                   /number= 34
FT   polyA_signal    5228..5233
FT                   /*tag=  ai
FT   misc_feature    5249..5253
FT                   /*tag=  aj
FT                   /note= "Poly adenylation cleavage site"
XX
PN   WO200050454-A1.
XX
PD   31-AUG-2000.
XX
PF   24-FEB-2000; 2000WO-US04732.
XX
PR   24-FEB-1999;   99US-0121461.
XX
PA   (TUFT ) TUFTS COLLEGE.
XX
PI   Soto AM,  Sonnenschein C,  Geck P,  Szelei J;
XX
DR   WPI; 2000-565451/52.
DR   P-PSDB; AAY94702.
XX
PT   New human androgen-induced tumor suppressor cDNA sequence termed
PT   'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
PT   diagnosis and treatment of prostate cancer -
XX
PS   Claim 1; Fig 1; 152pp; English.
XX
CC   This invention relates to a human androgen-induced tumour suppressor cDNA
CC   sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is located
CC   on chromosome 13 at position 13q12-13q. AS3 has a role in inhibiting cell
CC   proliferation and use as a marker for the efficient diagnosis and
CC   treatment of prostate cancer. The invention includes AS3 cDNA and protein
CC   sequences, a vector comprising the cDNA sequence, a host cell transfected
CC   with the expression vector, and a method for producing an AS3 polypeptide
CC   comprising culturing the transfected cells. AS3 has cytostatic activity,
CC   and acts to suppress cell proliferation. The AS3 gene is useful as a
```

```
CC   marker for the efficient diagnosis and treatment of prostate cancer. The
CC   AS3 nucleic acid molecule can be used as a source of antisense agents for
CC   sequence specific modulation of gene expression. The AS3 protein may be
CC   used in the treatment of disorders caused by aberrant modification or
CC   mutation of a gene encoding an AS3 protein, misregulation of the AS3 gene
CC   or aberrant post-translational modification of the AS3 protein. This
CC   sequence represents the human AS3 cDNA sequence.
XX
SQ   Sequence 5271 BP; 1782 A; 944 C; 1066 G; 1479 T; 0 other;

Alignment Scores:
Pred. No.:             0           Length:        5271
Score:                 7193.00     Matches:       1391
Percent Similarity:    100.00%     Conservative:  0
Best Local Similarity: 100.00%     Mismatches:    0
Query Match:           100.00%     Indels:        0
DB:                    21          Gaps:          0

US-09-512-581B-2 (1-1391) x AAA28051 (1-5271)

Qy      1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProProGlyValLys 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 125

Qy     21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValValLys 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    126 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 185

Qy     41 ThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAla 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    186 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 245

Qy     61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    246 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGGTAAAGATGTTCGCTTACTGGTA 305

Qy     81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    306 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 365

Qy    101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    366 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 425

Qy    121 LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSer 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    426 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 485

Qy    141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    486 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 545

Qy    161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    546 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 605

Qy    181 MetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    606 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 665

Qy    201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    666 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 725

Qy    221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    726 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCACTTTTTTTAATCAGGTT 785

Qy    241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeu 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    786 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 845

Qy    261 GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    846 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 905

Qy    281 LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    906 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 965

Qy    301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGly 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    966 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 1025

Qy    321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1026 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1085

Qy    341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1086 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1145

Qy    361 AspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1146 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1205

Qy    381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1206 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1265

Qy    401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAla 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1266 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1325

Qy    421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1326 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1385

Qy    441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1386 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1445

Qy    461 GlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1446 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1505

Qy    481 TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1506 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1565

Qy    501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1566 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1625

Qy    521 SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1626 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1685

Qy    541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArg 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1686 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1745

Qy    561 LysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysVal 580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1746 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1805

Qy    581 ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1806 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1865

Qy    601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1866 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 1925

Qy    621 LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyVal 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1926 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 1985

Qy    641 ProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHis 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1986 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 2045

Qy    661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2046 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2105

Qy    681 AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2106 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2165

Qy    701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLys 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2166 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2225

Qy    721 GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys 740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2226 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2285

Qy    741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2286 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2345

Qy    761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2346 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2405

Qy    781 PheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2406 TTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2465

Qy    801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerPro 820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2466 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2525

Qy    821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLys 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2526 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2585

Qy    841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2586 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2645

Qy    861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2646 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2705

Qy    881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2706 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2765

Qy    901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2766 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2825

Qy    921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet 940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2826 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2885

Qy    941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGln 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2886 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 2945

Qy    961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSer 980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2946 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 3005

Qy    981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3006 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3065

Qy   1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3066 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3125

Qy   1021 LeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle 1040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3126 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3185

Qy   1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3186 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3245

Qy   1061 MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3246 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3305

Qy   1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln 1100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3306 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3365

Qy   1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe 1120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3366 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3425

Qy   1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3426 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3485

Qy   1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3486 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3545

Qy   1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3546 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3605

Qy   1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3606 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3665

Qy   1201 ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThr 1220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3666 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3725

Qy   1221 ProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlu 1240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3726 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3785

Qy   1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3786 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3845

Qy   1261 GluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGlu 1280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3846 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3905

Qy   1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3906 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 3965

Qy   1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLysSerGly 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3966 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 4025

Qy   1321 ProProAlaProGluGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys 1340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4026 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4085

Qy   1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4086 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4145
```

```
Qy       1361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4146 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4205

Qy       1381 ThrProSerProSerGlnProLysLysAsnVal 1391
              |||||||||||||||||||||||||||||||||
Db       4206 ACGCCATCACCATCACAACCAAAAAAAAATGTG 4238

RESULT 2
AAA28052
ID   AAA28052 standard; cDNA; 5355 BP.
XX
AC   AAA28052;
XX
DT   01-DEC-2000  (first entry)
XX
DE   Human androgen shutoff gene 3 (AS3) cDNA sequence SEQ ID #4.
XX
KW   Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;
KW   chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;
KW   diagnosis; treatment; cytostatic; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   exon            85..130
FT                   /*tag= a
FT                   /number= 1
FT   exon            131..257
FT                   /*tag= b
FT                   /number= 2
FT   CDS             150..4325
FT                   /*tag= c
FT                   /product= "AS3"
FT                   /note= "Androgen shutoff gene 3 protein"
FT   exon            258..461
FT                   /*tag= d
FT                   /number= 3
FT   exon            462..548
FT                   /*tag= e
FT                   /number= 4
FT   exon            549..646
FT                   /*tag= f
FT                   /number= 5
FT   exon            647..773
FT                   /*tag= e
FT                   /number= 6
FT   exon            774..938
FT                   /*tag= f
FT                   /number= 7
FT   exon            939..995
FT                   /*tag= g
FT                   /number= 8
FT   exon            996..1111
FT                   /*tag= h
FT                   /number= 9
FT   exon            1112..1206
FT                   /*tag= i
FT                   /number= 10
FT   exon            1207..1352
FT                   /*tag= j
FT                   /number= 11
FT   exon            1353..1504
FT                   /*tag= k
FT                   /number= 12
FT   exon            1505..1618
FT                   /*tag= l
FT                   /number= 13
FT   exon            1619..1700
FT                   /*tag= m
FT                   /number= 14
FT   exon            1701..1749
FT                   /*tag= n
FT                   /number= 15
FT   exon            1750..1889
FT                   /*tag= o
FT                   /number= 16
FT   exon            1890..2005
FT                   /*tag= p
FT                   /number= 17
FT   exon            2006..2111
FT                   /*tag= q
FT                   /number= 18
FT   exon            2112..2272
FT                   /*tag= r
FT                   /number= 19
FT   exon            2273..2396
FT                   /*tag= s
FT                   /number= 20
FT   exon            2397..2555
FT                   /*tag= t
FT                   /number= 21
FT   exon            2556..2624
FT                   /*tag= u
FT                   /number= 22
FT   exon            2625..2761
FT                   /*tag= v
FT                   /number= 23
FT   exon            2672..2885
FT                   /*tag= w
FT                   /number= 24
FT   exon            2886..3090
FT                   /*tag= x
FT                   /number= 25
FT   exon            3091..3205
FT                   /*tag= y
FT                   /number= 26
FT   exon            3206..3338
FT                   /*tag= z
FT                   /number= 27
FT   exon            3339..3458
FT                   /*tag= aa
FT                   /number= 28
FT   exon            3459..3521
FT                   /*tag= ab
FT                   /number= 29
FT   exon            3522..3667
FT                   /*tag= ac
FT                   /number= 30
FT   exon            3668..3773
FT                   /*tag= ad
FT                   /number= 31
FT   exon            3774..4213
FT                   /*tag= ae
FT                   /number= 32
FT   exon            4214..4438
FT                   /*tag= af
FT                   /number= 33
FT   misc_signal     4314..4408
FT                   /*tag= ag
FT                   /note= "Destabilizing signal"
FT   exon            4439..5337
FT                   /*tag= ah
FT                   /number= 34
FT   polyA_signal    5312..5317
FT                   /*tag= ai
FT   misc_feature    5333..5337
FT                   /*tag= aj
FT                   /note= "Poly adenylation cleavage site"
XX
PN   WO200050454-A1.
XX
PD   31-AUG-2000.
XX
PF   24-FEB-2000; 2000WO-US04732.
XX
```

```
PR   24-FEB-1999;   99US-0121461.
XX
PA   (TUFT ) TUFTS COLLEGE.
XX
PI   Soto AM,  Sonnenschein C,  Geck P,  Szelei J;
XX
DR   WPI; 2000-565451/52.
DR   P-PSDB; AAY94702.
XX
PT   New human androgen-induced tumor suppressor cDNA sequence termed
PT   'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
PT   diagnosis and treatment of prostate cancer -
XX
PS   Example 4; Fig 6; 152pp; English.
XX
CC   This invention relates to a human androgen-induced tumour suppressor
CC   cDNA sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is
CC   located on chromosome 13 at position 13q12-13q. AS3 has a role in
CC   inhibiting cell proliferation and use as a marker for the efficient
CC   diagnosis and treatment of prostate cancer. The invention includes AS3
CC   cDNA and protein sequences, a vector comprising the cDNA sequence, a host
CC   cell transfected with the expression vector, and a method for producing
CC   an AS3 polypeptide comprising culturing the transfected cells. AS3 has
CC   cytostatic activity, and acts to suppress cell proliferation. The AS3
CC   gene is useful as a marker for the efficient diagnosis and treatment of
CC   prostate cancer. The AS3 nucleic acid molecule can be used as a source of
CC   antisense agents for sequence specific modulation of gene expression. The
CC   AS3 protein may be used in the treatment of disorders caused by aberrant
CC   modification or mutation of a gene encoding an AS3 protein, misregulation
CC   of the AS3 gene or aberrant post-translational modification of the AS3
CC   protein. This sequence represents the human AS3 cDNA sequence with an
CC   additional 84 nucleotides in the 5' untranslated region (5' UTR) when
CC   compared with the claimed AS3 cDNA sequence AAA28051.
XX
SQ   Sequence 5355 BP; 1798 A; 957 C; 1115 G; 1485 T; 0 other;

Alignment Scores:
Pred. No.:               0              Length:         5355
Score:                   7193.00        Matches:        1391
Percent Similarity:      100.00%        Conservative:   0
Best Local Similarity:   100.00%        Mismatches:     0
Query Match:             100.00%        Indels:         0
DB:                      21             Gaps:           0

US-09-512-581B-2 (1-1391) x AAA28052 (1-5355)

Qy          1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProProGlyValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        150 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 209

Qy         21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValValLys 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        210 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 269

Qy         41 ThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        270 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 329

Qy         61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        330 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGGTAAAGATGTTCGCTTACTGGTA 389

Qy         81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        390 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 449

Qy        101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        450 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 509

Qy        121 LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        510 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 569

Qy        141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        570 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 629

Qy        161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        630 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 689

Qy        181 MetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 749

Qy        201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        750 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 809

Qy        221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        810 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCACTTTTTTTAATCAGGTT 869

Qy        241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        870 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 929

Qy        261 GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        930 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 989

Qy        281 LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        990 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 1049

Qy        301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGly 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1050 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 1109

Qy        321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1110 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1169

Qy        341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1170 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1229

Qy        361 AspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1230 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1289

Qy        381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1290 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1349

Qy        401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAla 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1350 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1409

Qy        421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1410 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1469

Qy        441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1470 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1529

Qy        461 GlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1530 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1589

Qy        481 TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1590 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1649
```

```
Qy        501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1650 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1709

Qy        521 SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1710 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1769

Qy        541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArg 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1770 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1829

Qy        561 LysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysVal 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1830 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1889

Qy        581 ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1890 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1949

Qy        601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1950 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 2009

Qy        621 LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyVal 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2010 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 2069

Qy        641 ProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHis 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2070 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 2129

Qy        661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2130 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2189

Qy        681 AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2190 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2249

Qy        701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLys 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2250 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2309

Qy        721 GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2310 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2369

Qy        741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2370 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2429

Qy        761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2430 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2489

Qy        781 PheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2490 TTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2549

Qy        801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerPro 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2550 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2609

Qy        821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLys 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2610 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2669

Qy        841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2670 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2729

Qy        861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2730 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2789

Qy        881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2790 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2849

Qy        901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2850 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2909

Qy        921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet 940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2910 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2969

Qy        941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGln 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2970 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 3029

Qy        961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSer 980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3030 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 3089

Qy        981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3090 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3149

Qy       1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3150 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3209

Qy       1021 LeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle 1040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3210 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3269

Qy       1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3270 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3329

Qy       1061 MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3330 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3389

Qy       1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln 1100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3390 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3449

Qy       1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe 1120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3450 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3509

Qy       1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3510 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3569

Qy       1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3570 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3629

Qy       1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3630 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3689

Qy       1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3690 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3749

Qy       1201 ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThr 1220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3750 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3809

Qy       1221 ProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlu 1240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       3810 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3869

Qy       1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3870 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3929

Qy       1261 GluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGlu 1280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3930 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3989

Qy       1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3990 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 4049

Qy       1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLysSerGly 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4050 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 4109

Qy       1321 ProProAlaProGluGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys 1340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4110 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4169

Qy       1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4170 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4229

Qy       1361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4230 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4289

Qy       1381 ThrProSerProSerGlnProLysLysAsnVal 1391
              |||||||||||||||||||||||||||||||||
Db       4290 ACGCCATCACCATCACAACCAAAAAAAAATGTG 4322
```

RESULT 3
ABV22430
```
ID   ABV22430 standard; cDNA; 7473 BP.
XX
AC   ABV22430;
XX
DT   13-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 22421.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer  -
XX
PS   Claim 1; Page 3901-3902; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ   Sequence 7473 BP; 2458 A; 1243 C; 1480 G; 2285 T; 7 other;

Alignment Scores:
Pred. No.:              0           Length:        7473
Score:                  7155.00     Matches:       1387
Percent Similarity:     99.71%      Conservative:  0
Best Local Similarity:  99.71%      Mismatches:    4
Query Match:            99.47%      Indels:        0
DB:                     23          Gaps:          0

US-09-512-581B-2 (1-1391) x ABV22430 (1-7473)

Qy          1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProProGlyValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        158 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 217

Qy         21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValValLys 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        218 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 277

Qy         41 ThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        278 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 337

Qy         61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
              |||||||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db        338 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGATAAAGATGTTCGCTTACTGGTA 397

Qy         81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        398 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 457

Qy        101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        458 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 517

Qy        121 LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        518 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 577

Qy        141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        578 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 637

Qy        161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        638 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 697

Qy        181 MetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        698 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 757

Qy        201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        758 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 817
```

```
Qy        221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
              ||||||||||||||||||||||||||||||||||||||||||||     |||||||||||||||
Db        818 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTTTAATCAGGTT 877

Qy        241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        878 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 937

Qy        261 GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        938 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 997

Qy        281 LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        998 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 1057

Qy        301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGly 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1058 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 1117

Qy        321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1118 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1177

Qy        341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1178 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1237

Qy        361 AspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1238 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1297

Qy        381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1298 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1357

Qy        401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAla 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1358 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1417

Qy        421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1418 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1477

Qy        441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1478 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1537

Qy        461 GlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1538 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1597

Qy        481 TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1598 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1657

Qy        501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1658 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1717

Qy        521 SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1718 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1777

Qy        541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArg 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1778 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1837

Qy        561 LysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysVal 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1838 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1897

Qy        581 ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1898 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1957

Qy        601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1958 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 2017

Qy        621 LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyVal 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2018 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 2077

Qy        641 ProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHis 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2078 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 2137

Qy        661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2138 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2197

Qy        681 AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2198 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2257

Qy        701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLys 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2258 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2317

Qy        721 GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2318 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2377

Qy        741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2378 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2437

Qy        761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2438 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2497

Qy        781 PheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
              |||||||||||     ||||||     |||||||||||||||||||||||||||||||||
Db       2498 TTTGCTGCTCCTTTGAAATCTTTGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2557

Qy        801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerPro 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2558 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2617

Qy        821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLys 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2618 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2677

Qy        841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2678 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2737

Qy        861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2738 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2797

Qy        881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2798 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2857

Qy        901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2858 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2917

Qy        921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet 940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2918 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2977

Qy        941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGln 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2978 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 3037

Qy        961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSer 980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3038 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 3097

Qy        981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3098 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3157

Qy       1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3158 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3217

Qy       1021 LeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle 1040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3218 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3277

Qy       1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3278 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3337

Qy       1061 MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3338 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3397

Qy       1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln 1100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3398 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3457

Qy       1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe 1120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3458 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3517

Qy       1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3518 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3577

Qy       1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3578 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3637

Qy       1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3638 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3697

Qy       1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3698 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3757

Qy       1201 ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThr 1220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3758 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3817

Qy       1221 ProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlu 1240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3818 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3877

Qy       1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3878 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3937

Qy       1261 GluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGlu 1280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3938 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3997

Qy       1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3998 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 4057

Qy       1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLysSerGly 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4058 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 4117

Qy       1321 ProProAlaProGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys 1340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4118 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4177

Qy       1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4178 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4237

Qy       1361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4238 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4297

Qy       1381 ThrProSerProSerGlnProLysLysAsnVal 1391
              |||||||||||||||||||||||||||||||||
Db       4298 ACGCCATCACCATCACAACCAAAAAAAAATGTC 4330
```

```
RESULT 4
ABV28244
ID   ABV28244 standard; cDNA; 7473 BP.
XX
AC   ABV28244;
XX
DT   16-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 28235.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer  -
XX
PS   Claim 1; Page 5869-5870; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
```

```
SQ   Sequence 7473 BP; 2458 A; 1244 C; 1479 G; 2285 T; 7 other;

Alignment Scores:
Pred. No.:              0           Length:        7473
Score:                  7149.00     Matches:       1386
Percent Similarity:     99.64%      Conservative:  0
Best Local Similarity:  99.64%      Mismatches:    5
Query Match:            99.39%      Indels:        0
DB:                     23          Gaps:          0

US-09-512-581B-2 (1-1391) x ABV28244 (1-7473)

Qy        1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProProGlyValLys 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      158 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 217

Qy       21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValValLys 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      218 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 277

Qy       41 ThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAla 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      278 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 337

Qy       61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
            ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db      338 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGATAAAGATGTTCGCTTACTGGTA 397

Qy       81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      398 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 457

Qy      101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      458 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 517

Qy      121 LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSer 140
            ||||||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db      518 AAGAGCCCACAATTCAATAGCTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 577

Qy      141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      578 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 637

Qy      161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      638 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 697

Qy      181 MetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      698 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 757

Qy      201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      758 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 817

Qy      221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
            ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db      818 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTTTTAATCAGGTT 877

Qy      241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeu 260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      878 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 937

Qy      261 GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      938 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 997

Qy      281 LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      998 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 1057

Qy      301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGly 320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1058 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 1117

Qy      321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1118 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1177

Qy      341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1178 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1237

Qy      361 AspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys 380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1238 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1297

Qy      381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1298 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1357

Qy      401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAla 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1358 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1417

Qy      421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu 440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1418 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1477

Qy      441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1478 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1537

Qy      461 GlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1538 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1597

Qy      481 TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1598 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1657

Qy      501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1658 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1717

Qy      521 SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1718 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1777

Qy      541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArg 560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1778 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1837

Qy      561 LysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysVal 580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1838 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1897

Qy      581 ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1898 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1957

Qy      601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla 620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1958 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 2017

Qy      621 LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyVal 640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2018 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 2077

Qy      641 ProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHis 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2078 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 2137

Qy      661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      2138 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2197

Qy       681 AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2198 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2257

Qy       701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLys 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2258 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2317

Qy       721 GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys 740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2318 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2377

Qy       741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2378 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2437

Qy       761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2438 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2497

Qy       781 PheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
             ||||||||||||   ||||||   ||||||||||||||||||||||||||||||||||||
Db      2498 TTTGCTGCTCCTTTGAAATCTTTGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2557

Qy       801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerPro 820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2558 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2617

Qy       821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLys 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2618 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2677

Qy       841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2678 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2737

Qy       861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2738 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2797

Qy       881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2798 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2857

Qy       901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2858 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2917

Qy       921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet 940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2918 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2977

Qy       941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGln 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2978 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 3037

Qy       961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSer 980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3038 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 3097

Qy       981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3098 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3157

Qy      1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3158 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3217

Qy      1021 LeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle 1040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3218 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3277

Qy      1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3278 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3337

Qy      1061 MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3338 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3397

Qy      1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln 1100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3398 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3457

Qy      1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe 1120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3458 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3517

Qy      1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3518 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3577

Qy      1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3578 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3637

Qy      1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3638 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3697

Qy      1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3698 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3757

Qy      1201 ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThr 1220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3758 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3817

Qy      1221 ProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlu 1240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3818 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3877

Qy      1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3878 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3937

Qy      1261 GluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGlu 1280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3938 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3997

Qy      1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3998 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 4057

Qy      1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLysSerGly 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4058 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 4117

Qy      1321 ProProAlaProGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys 1340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4118 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4177

Qy      1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4178 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4237

Qy      1361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4238 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4297

Qy      1381 ThrProSerProSerGlnProLysLysAsnVal 1391
             |||||||||||||||||||||||||||||||||
Db      4298 ACGCCATCACCATCACAACCAAAAAAAAATGTC 4330
```

```
RESULT 5
ABV25469
ID   ABV25469 standard; cDNA; 7473 BP.
XX
AC   ABV25469;
XX
DT   16-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 25460.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer  -
XX
PS   Claim 1; Page 5055-5057; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ   Sequence 7473 BP; 2457 A; 1243 C; 1481 G; 2285 T; 7 other;

Alignment Scores:
Pred. No.:              0              Length:        7473
Score:                  7148.00        Matches:       1386
Percent Similarity:     99.64%         Conservative:  0
Best Local Similarity:  99.64%         Mismatches:    5
Query Match:            99.37%         Indels:        0
DB:                     23             Gaps:          0

US-09-512-581B-2 (1-1391) x ABV25469 (1-7473)

Qy          1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProProGlyValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        158 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 217

Qy         21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValValLys 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        218 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 277

Qy         41 ThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        278 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 337

Qy         61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
              ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db        338 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGATAAAGATGTTCGCTTACTGGTA 397

Qy         81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        398 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 457

Qy        101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        458 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 517

Qy        121 LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        518 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 577

Qy        141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        578 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 637

Qy        161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        638 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 697

Qy        181 MetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        698 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 757

Qy        201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        758 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 817

Qy        221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
              ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db        818 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTTTTAATCAGGTT 877

Qy        241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        878 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 937

Qy        261 GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        938 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 997

Qy        281 LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        998 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 1057

Qy        301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGly 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1058 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 1117

Qy        321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1118 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1177

Qy        341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1178 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1237

Qy        361 AspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1238 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1297

Qy        381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1298 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1357

Qy        401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAla 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1358 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1417

Qy        421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1418 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1477

Qy        441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1478 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1537

Qy        461 GlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1538 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1597

Qy        481 TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1598 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1657

Qy        501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1658 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1717

Qy        521 SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1718 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1777

Qy        541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArg 560
              |||||||||||||||||||||||||||||||||||||     ||||||||||||||||||
Db       1778 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGGAGATGATGAGAAAATAAGA 1837

Qy        561 LysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysVal 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1838 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1897

Qy        581 ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1898 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1957

Qy        601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1958 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 2017

Qy        621 LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyVal 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2018 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 2077

Qy        641 ProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHis 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2078 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 2137

Qy        661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2138 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2197

Qy        681 AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2198 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2257

Qy        701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLys 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2258 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2317

Qy        721 GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2318 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2377

Qy        741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2378 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2437

Qy        761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2438 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2497

Qy        781 PheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
              ||||||||||||   ||||||    |||||||||||||||||||||||||||||||||||
Db       2498 TTTGCTGCTCCTTTGAAATCTTTGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2557

Qy        801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerPro 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2558 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2617

Qy        821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLys 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2618 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2677

Qy        841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2678 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2737

Qy        861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2738 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2797

Qy        881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2798 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2857

Qy        901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2858 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2917

Qy        921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet 940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2918 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2977

Qy        941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGln 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2978 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 3037

Qy        961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSer 980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3038 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 3097

Qy        981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3098 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3157

Qy       1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3158 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3217

Qy       1021 LeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle 1040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3218 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3277

Qy       1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3278 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3337

Qy       1061 MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3338 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3397

Qy       1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln 1100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3398 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3457

Qy       1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe 1120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3458 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3517
```

```
Qy       1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3518 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3577

Qy       1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3578 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3637

Qy       1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3638 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3697

Qy       1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3698 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3757

Qy       1201 ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThr 1220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3758 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3817

Qy       1221 ProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlu 1240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3818 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3877

Qy       1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3878 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3937

Qy       1261 GluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGlu 1280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3938 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3997

Qy       1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3998 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 4057

Qy       1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLysSerGly 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4058 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 4117

Qy       1321 ProProAlaProGluGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys 1340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4118 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4177

Qy       1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4178 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4237

Qy       1361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4238 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4297

Qy       1381 ThrProSerProSerGlnProLysLysAsnVal 1391
              |||||||||||||||||||||||||||||||||
Db       4298 ACGCCATCACCATCACAACCAAAAAAAAATGTC 4330
```

```
RESULT 6
ABX14056
ID   ABX14056 standard; cDNA; 6744 BP.
XX
AC   ABX14056;
XX
DT   24-FEB-2003  (first entry)
XX
DE   cDNA encoding human SCC-112 tumour supressor gene.
XX
KW   ss; gene; human; chromosome 4p14; SCC-112; cancer; apoptosis; allodynia;
KW   degenerative disorder; metastasis inhibition; breast cancer; causalgia;
KW   kidney cancer; bladder cancer; pancreatic cancer; colon cancer;
KW   squamous cell carcinoma; head trauma; spinal cord injury; herpes zoster;
KW   global and focal ischaemic and haemorrhagic stroke; epilepsy; neuralgia;
KW   hypoxia-induced nerve cell damage; anxiety; diabetes mellitus;
KW   cardiac arrest; spinal cord lesion; stomach cancer; lung cancer;
KW   neonatal distress; Alzheimer's disease; uterine cancer; colon cancer;
KW   multiple sclerosis; phantom limb pain; hyperalgesia; Down's syndrome;
KW   Huntington's disease; Parkinson's disease; Korsakoff's syndrome;
KW   amyotrophic lateral sclerosis; cell survival; cell proliferation;
KW   tumour supressor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             232..4125
FT                   /*tag=  a
FT                   /product= "Human SCC-112"
FT   polyA_signal    6721..6726
FT                   /*tag=  b
FT                   /standard_name= "PolyA signal"
XX
PN   WO200281641-A2.
XX
PD   17-OCT-2002.
XX
PF   08-APR-2002; 2002WO-US10850.
XX
PR   06-APR-2001; 2001US-281780P.
XX
PA   (GEOU ) UNIV GEORGETOWN.
XX
PI   Kasid UN,  Kumar D,  Ahmad I;
XX
DR   WPI; 2003-103330/09.
DR   P-PSDB; ABG72803.
XX
PT   New isolated SSC (undefined) tumor suppressor polypeptides and
PT   polynucleotides, useful for diagnosing, preventing or treating cancer
PT   or degenerative disease, e.g. Alzheimer's Disease, Huntington's
PT   disease, or multiple sclerosis  -
XX
PS   Claim 1; Fig 1B; 83pp; English.
XX
CC   The invention relates to a new isolated polypeptide SCC-112. The SCC-112
CC   polypeptides and polynucleotides are useful for diagnosing, preventing or
CC   treating cancer (e.g. breast cancer, kidney cancer, bladder cancer,
CC   pancreatic cancer, colon cancer, squamous cell carcinoma, uterine
CC   cancer, stomach cancer, colon cancer, lung cancer); or degenerative
CC   disease or disorder (e.g. global and focal ischaemic and haemorrhagic
CC   stroke, head trauma, spinal cord injury, hypoxia-induced nerve cell
CC   damage, nerve cell damage caused by cardiac arrest or neonatal distress,
CC   epilepsy, anxiety, diabetes mellitus, multiple sclerosis, phantom limb
CC   pain, causalgia, neuralgias, herpes zoster, spinal cord lesions,
CC   hyperalgesia, allodynia, Alzheimer's disease, Huntington's disease,
CC   Parkinson's disease, amyotrophic lateral sclerosis, Down's syndrome and
CC   Korsakoff's syndrome). The polypeptides and polynucleotides are also
CC   useful for inducing apoptosis in cancer cells, increasing survival or
CC   proliferation of a cell, or inhibiting cancer cell proliferation and/or
CC   metastasis in a cancer patient. The polynucleotides can be used as probes
CC   to detect complementary nucleotide sequences, or as primers to obtain
CC   additional copies of the polynucleotides. SSC-112 may also be used for
CC   identifying drugs for treatment of cancers. The present sequence
CC   represents cDNA encoding the human SCC-112 tumour supressor gene which
CC   is located on chromosome 4p14.
XX
SQ   Sequence 6744 BP; 2156 A; 1243 C; 1400 G; 1945 T; 0 other;

Alignment Scores:
Pred. No.:              0          Length:        6744
Score:                  4760.50    Matches:       935
Percent Similarity:     80.29%     Conservative:  173
Best Local Similarity:  67.75%     Mismatches:    210
Query Match:            66.18%     Indels:        63
DB:                     25         Gaps:          15

US-09-512-581B-2 (1-1391) x ABX14056 (1-6744)

Qy          7 ArgThrAsnAspGlyLysIleThrTyrProProGlyValLysGluIleSerAspLysIle 26
```

```
Db    160 CGCACACACTCCCGGAAGATCGCTTACCCTCCGGGGGTAAAAGAGATCACCGACAAGATC 219

Qy     27 SerLysGluGluMetValArgArgLeuLysMetValValLysThrPheMetAspMetAsp 46
Db    220 ACCACGGACGAGATGATCAAACGCCTGAAGATGGTAGTGAAAACCTTTATGGATATGGAT 279

Qy     47 GlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
Db    280 CAGGACTCAGAAGATGAAAAACAGCAGTATCTCCCACTAGCCTTGCATCTTGCATCTGAA 339

Qy     67 PhePheLeuLysHisProGlyLysAspValArgLeuLeuValAlaCysCysLeuAlaAsp 86
Db    340 TTCTTCCTCAGGAACCCCAATAAAGATGTGCGTCTCCTTGTAGCATGTTGTTTGGCTGAT 399

Qy     87 IlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLysLeuLysAspIle 106
Db    400 ATCTTTCGTATCTATGCCCCAGAAGCTCCATATACTTCCCATGATAAACTTAAGGACATA 459

Qy    107 PheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAsn 126
Db    460 TTTTTGTTTATTACCAGACAATTAAAAGGTTTGGAGGATACAAAGAGTCCACAGTTTAAT 519

Qy    127 ArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsnIleCysPheGlu 146
Db    520 AGATACTTTTATTTATTAGAGAATTTAGCTTGGGTTAAATCATATAACATCTGCTTTGAA 579

Qy    147 LeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsn 166
Db    580 TTGGAAGATTGCAATGAAATTTTTATTCAGCTTTTTAGAACTCTCTTCTCAGTGATCAAC 639

Qy    167 AsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSerSerIleIleCys 186
Db    640 AATAGCCACAATAAGAAGGTACAAATGCACATGCTAGATTTGATGAGTTCTATCATCATG 699

Qy    187 GluGlyAspThrValSerGlnGluLeuLeuAspThrValLeuValAsnLeuValProAla 206
Db    700 GAAGGTGATGGAGTTACTCAAGAATTATTGGGCTCCATTCTTATTAACCTCATTCCTGCA 759

Qy    207 HisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAla 226
Db    760 CATAAGAACTTAAATAAACAGTCCTTTGACCTTGCAAAAGTGCTATTGAAAAGAACAGTC 819

Qy    227 GlnAlaIleGluProTyrIleThrThrPhePheAsnGlnValLeuMetLeuGlyLysThr 246
Db    820 CAGACTATTGAGGCATGCATTGCCAATTTTTTCAATCAAGTCCTGGTGCTGGGAAGATCA 879

Qy    247 SerIleSerAspLeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAsp 266
Db    880 TCAGTAAGTGATTTGTCAGAACATGTATTTGATCTGATTCAGGAACTTTTTGCTATAGAT 939

Qy    267 SerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerAsnAspAsn 286
Db    940 CCTCATTTATTATTATCCGTCATGCCACAGCTTGAATTCAAACTAAAGAGCAATGATGGA 999

Qy    287 GluGluArgLeuGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGlu 306
Db   1000 GAAGAGCGATTAGCTGTTGTTCGACTTCTAGCTAAATTGTTTGGCTCCAAAGATTCTGAT 1059

Qy    307 LeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHis 326
Db   1060 TTGGCAACACAGAATCGTCCTCTTTGGCAATGTTTTCTTGGACGATTTAATGATATTCAT 1119

Qy    327 ValProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAsp 346
Db   1120 GTTCCTGTGAGATTAGAAAGTGTGAAATTTGCCAGTCATTGTTTAATGAATCACCCAGAT 1179

Qy    347 LeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIle 366
Db   1180 TTAGCGAAGGATCTCACAGAATATTTAAAGGTTAGATCACATGATCCAGAAGAAGCTATT 1239

Qy    367 ArgHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeuLeuValAsn 386
Db   1240 CGTCATGATGTCATTGTTACTATAATAACAGCTGCCAAGAGGGACCTGGCCTTAGTAAAT 1299

Qy    387 AspHisLeuLeuAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLys 406
Db   1300 GATCAGCTGCTTGGCTTTGTAAGGGAAAGAACACTGGATAAACGGTGGCGAGTAAGAAAA 1359

Qy    407 GluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGly 426
Db   1360 GAAGCTATGATGGGTCTGGCTCAGCTTTATAAGAAATACTGTCTTCATGGTGAAGCAGGA 1419

Qy    427 LysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGln 446
Db   1420 AAGGAAGCTGCAGAGAAAGTCAGCTGGATAAAGGACAAACTTCTGCATATTTATTATCAG 1479

Qy    447 AsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHis 466
Db   1480 AACAGCATTGACGACAAACTGTTGGTAGAGAAAATCTTTGCTCAGTATCTTGTCCCCCAC 1539

Qy    467 AsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeu 486
Db   1540 AACCTGGAAACAGAAGAGAGAATGAAATGCTTATATTACTTATATGCTAGTTTGGATCCA 1599

Qy    487 AsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnVal 506
Db   1600 AATGCTGTAAAAGCTCTCAACGAAATGTGGAAGTGTCAGAACATGCTTCGGAGCCATGTA 1659

Qy    507 LysAspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePhe 526
Db   1660 CGCGAACTATTGGATTTGCACAAGCAGCCTACATCAGAGGCTAACTGTTCTGCCATGTTT 1719

Qy    527 SerLysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMet 546
Db   1720 GGAAAACTGATGACCATAGCAAAGAATTTGCCTGACCCCGGGAAAGCACAAGATTTTGTG 1779

Qy    547 LysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeu 566
Db   1780 AAGAAATTTAACCAGGTTCTCGGCGATGATGAGAAACTTCGGTCTCAGTTGGAGTTATTA 1839

Qy    567 ValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLys 586
Db   1840 ATTAGCCCAACCTGTTCTTGCAAACAAGCAGATATTTGTGTGAGAGAAATAGCCCGGAAA 1899

Qy    587 LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGlu 606
Db   1900 CTTGCAAATCCTAAGCAACCAACAAATCCTTTTCTAGAGATGGTCAAATTTCTGTTGGAA 1959

Qy    607 ArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsn 626
Db   1960 AGAATCGCACCTGTGCACATTGATTCAGAAGCCATAAGTGCACTAGTGAAATTGATGAAT 2019

Qy    627 LysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIle 646
Db   2020 AAGTCAATAGAGGGGACAGCAGATGATGAAGAGGAGGGTGTAAGTCCAGATACAGCTATC 2079

Qy    647 ArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSer 666
Db   2080 CGTTCAGGACTTGAACTTCTTAAGGTTCTGTCTTTTACACATCCTACCTCGTTCCACTCT 2139

Qy    667 AlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGlu 686
Db   2140 GCAGAGACATATGAGTCCTTGTTACAGTGCCTAAGAATGGAGGATGACAAGGTAGCAGAA 2199

Qy    687 AlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHisIle 706
Db   2200 GCTGCTATTCAAATTTTTAGAAATACAGGTCACAAAATAGAAACAGACCTTCCCCAGATA 2259

Qy    707 ArgSerAlaLeuLeuProValLeuHisHisLysSerLysLysGlyProProArgGlnAla 726
Db   2260 CGATCGACCTTAATTCCCATTTTACATCAAAAAGCAAAGAGGGGTACTCCACACCAAGCA 2319

Qy    727 LysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGln 746
Db   2320 AAACAGGCTGTGCACTGTATACACGCCATATTCACAAATAAAGAAGTCCAGCTTGCACAG 2379
```

```
Qy        747 IlePheGluProLeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleThrPro 766
              ||||||||||||||     :::||||||:::               |||   ||||||||||||
Db       2380 ATTTTTGAGCCACTCAGTAGGAGTCTGAATGCTGATGTGCCAGAACAACTTATAACTCCA 2439

Qy        767 LeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLys 786
              |||||:::::||||||||:::::||||||||||||||||||||||:::|||   |||
Db       2440 TTAGTTTCATTGGGCCACATTTCTATGTTAGCACCAGATCAGTTTGCTTCCCCAATGAAA 2499

Qy        787 SerTrpValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLys 806
              |||  ||||||  ||||||||||||||||||||||||||||||||||         |||:::
Db       2500 TCTGTAGTAGCAAATTTTATTGTGAAAGATCTGCTAATGAATGACAGGTCAACAGGTGAA 2559

Qy        807 LysThrThrLysLeuTrpValProAspGluGluValSerProGluThrMetValLysIle 826
              |||      |||||||||  |||||||||||||||||||||||||   :::   |||:::
Db       2560 AAGAATGGAAAACTGTGGTCTCCAGATGAAGAGGTTTCCCCTGAAGTACTAGCAAAGGTA 2619

Qy        827 GlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSer 846
              ||||||||||||:::::|||||||||||||||||||||||||||||||||   |||||||||
Db       2620 CAGGCAATTAAACTTCTGGTAAGGTGGCTGTTGGGTATGAAAAACAACCAGTCTAAATCT 2679

Qy        847 GlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGlu 866
                       ||||||||||||||||||:::   :::|||   |||:::|||||||||||||||
Db       2680 GCCAATTCAACCCTTCGGTTATTATCAGCGATGTTGGTTAGTGAGGGTGACCTGACAGAG 2739

Qy        867 GlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIle 886
              |||   :::|||||||||  ||||||||||||||||||||||||||||||||||||||||
Db       2740 CAAAAGAGGATCAGTAAATCTGATATGTCTCGCTTGCGATTAGCTGCTGGTAGTGCCATA 2799

Qy        887 ValLysLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGluGlnTyrGlnLeu 906
              :::|||||||||||||||||||||||||||||||||||||||||   ||||||:::||||||
Db       2800 ATGAAGCTTGCTCAGGAACCTTGTTACCATGAAATTATTACCCCAGAACAGTTTCAGCTC 2859

Qy        907 CysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeu 926
              |||||||||   |||||||||||||||||||||||||||||||||||:::||||||||||||
Db       2860 TGTGCACTTGTTATTAATGATGAGTGTTACCAAGTAAGGCAGATATTTGCTCAGAAGCTG 2919

Qy        927 HisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCys 946
              ||||||   |||   :::|||   ||||||||||||||||||||||||     |||||||||
Db       2920 CATAAGGCACTTGTGAAGTTACTGCTCCCATTGGAGTATATGGCGATCTTTGCCTTGTGT 2979

Qy        947 AlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIle 966
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db       2980 GCCAAAGATCCTGTGAAGGAGAGAAGAGCACACGCACGACAATGTTTACTGAAAAATATC 3039

Qy        967 AsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeu 986
              ::::::|||||||||||||:::|||||||:::               :::||||||||||||||||
Db       3040 AGTATACGCAGGGAATACATTAAGCAGAATCCTATGGCTACTGAGAAATTATTATCACTG 3099

Qy        987 LeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrVal 1006
              |||||||||||||||||||||||||||   |||||||||||||||||||||||||||:::
Db       3100 TTGCCTGAATATGTAGTTCCATACATGATTCACCTGCTAGCCCATGATCCAGATTTACA 3159

Qy       1007 LysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValLeuGlu 1026
              :::   |||||::::::||||||:::|||:::||||||||||||||||||:::||||||
Db       3160 AGATCACAAGATGTTGATCAGCTTCGTGATATCAAAGAGTGCCTATGGTTCATGCTTGAA 3219

Qy       1027 IleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsn 1046
              :::||||||   |||||||||||||||||||||||||||||:::::::||||||   ||||||
Db       3220 GTTTTAATGACAAAGAATGAAAACAATAGCCATGCCTTTATGAAGAAGATGGCAGAGAAC 3279

Qy       1047 IleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyr 1066
              ||||||   |||:::||||||||||   ||||||::::::|||   ||||||||||||||||
Db       3280 ATCAAGTTAACCAGAGATGCCCAGTCTCCAGATGAATCCAAGACAAATGAAAAACTGTAT 3339

Qy       1067 ThrValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSerLeuGlu 1086
              |||||||||||||||||:::   :::|||   |||||||||          :::   :::
Db       3340 ACAGTATGTGATGTGGCTCTCTGTGTTATAAATAGTAAAAGTGCTTTGTGCAATGCAGAT 3399

Qy       1087 SerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsnPheSer 1106
              ||||||||||||||||||||||||       :::|||||||||||||||:::|||:::|||
Db       3400 TCACCAAAGGACCCAGTCCTCCCAATGAAATTTTTACACAACCTGAAAAGGACTTCTGT 3459

Qy       1107 AsnThrLysAsnTyrLeuProProGluMetLysSerPhePheThrProGlyLysProLys 1126
              |||   |||:::|||:::         |||   :::               ||||||||||||
Db       3460 AACGATAAGAGTTATATTTCAGAAGAGACAAGAGTACTTCTGTTAACAGGAAAGCCAAAG 3519

Qy       1127 ThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGln 1146
                    |||||||||||||||||||||||||||||||:::   |||::::::
Db       3520 CCTGCTGGAGTACTAGGTGCAGTAAATAAGCCTTTATCAGCAACGGGAAGGAAACCCTAT 3579

Qy       1147 ThrLysSerSerArgMetGluThrValSer------AsnAlaSerSerSerAsnPro 1164
              :::|||:::         ||||||   |||            |||   :::|||      ||||||
Db       3580 GTTAGAAGCACTGGCACTGAGACTGGAAGCAATATTAATGTAAATTCAGAGCTGAACCCT 3639

Qy       1165 SerSerProGlyArgIleLysGlyArg---LeuAspSerSerGluMetAspHisSerGlu 1183
              |||:::      |||   :::   :::      :::::::::|||         ||||||
Db       3640 TCAACCGGAAATCGATCAAGGGAACAGAGTTCAGAGGCAGCAGAAACTGGAGTTAGTGAA 3699

Qy       1184 AsnGluAsp------TyrThrMetSerSerProLeuProGlyLysLysSerAspLysArg 1201
              ||||||:::                  :::   |||         |||   |||      |||
Db       3700 AATGAAGAGAACCCTGTGAGGATTATTTCAGTCACACCTGTAAAGAATATTGAC------ 3753

Qy       1202 AspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThrPro 1221
                                                                         |||
Db       3754 ------------------------------------------------------CCA 3756

Qy       1222 ValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGluGln 1241
              |||      :::|||      :::         |||      |||               |||
Db       3757 GTAAAGAATAAGGAA------ATTAATTCTGATCAGGCTACC--------------CAG 3795

Qy       1242 LysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGlu 1261
                          |||   |||   :::||||||      ||||||:::   ::::::
Db       3796 GGCAACATCAGCAGTGACCGAGGAAAGAAAAGAACAGTAACAGCAGCTGGTGCAGAGAAT 3855

Qy       1262 ---GlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGlu 1280
                 ||||||      :::||||||                              ||||||
Db       3856 ATCCAACAAAAAACAGATGAGAAA-------------------------GTAGATGAA 3888

Qy       1281 GlnAsnSerPro---ProLysLysGlyLysArgGlyArgProProLysProLeuGly--- 1298
                       |||   |||   |||   :::|||||||||      ||||||
Db       3889 TCGGGACCTCCCGCCCCTTCCAAACCCAGGAGAGGACGTCGACCCAAGTCTGAATCTCAG 3948

Qy       1299 GlyGlyThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLys 1318
              |||      |||   :::         |||         ||||||   |||:::
Db       3949 GGCAATGCTACCAAAAATGATGATCTAAATAAACCTATTAACAAGGGAAGGAAGAGAGCT 4008

Qy       1319 Ser---GlyProProAlaProGluGluGluGlu------------------------ 1328
              :::   |||      :::|||         |||
Db       4009 GCAGTGGGTCAGGAGAGCCCTGGGGGTTTGGAAGCAGGTAATGCCAAAGCACCCAAACTG 4068

Qy       1329 GluGluGluArgGlnSerGlyAsnThrGluGlnLys-----------SerLysSerLys 1344
              ::::::      |||   |||:::|||   :::|||            :::|||   |||
Db       4069 CAAGATTT-AGCCAAAAAGGCAGCACCAGCAGAAAGACAAATTGACTTACAAAGGTAAAA 4127

Qy       1345 GlnHis------ArgValSerArgArgAlaGlnGlnArgAlaGluSerProGluSerSer 1362
                 |||      |||            :::|||:::|||:::   ::::::|||   |||:::
Db       4128 ATGCATTTGCAAAGGGAGAAAATGAAGGCCAAACAGAAG---CAGGCTCCAGCTTCTGCA 4184
```

RESULT 7
ABQ99271
ID   ABQ99271 standard; cDNA; 5469 BP.
XX
AC   ABQ99271;
XX
DT   25-FEB-2003 (first entry)
XX
DE   Human coding sequence SEQ ID 4.
XX
KW   Human; expressed sequence tag; EST; chromosome 4;
KW   haematopoietic disorder; central nervous system disease; viral infection;
KW   peripheral nervous system disease; non-healing wound; infectious disease;
KW   immune deficiency; immune disorder; bacterial infection; allergy; cancer;
KW   fungal infection; autoimmune disorder; coagulation disorder; nootropic;
KW   antiallergic; antiinflammatory; immunosuppressive; neuroprotective;
KW   cytostatic; haemostatic; virucide; antibacterial; fungicide;

```
KW   immunostimulant; cerebroprotective; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200259260-A2.
XX
PD   01-AUG-2002.
XX
PF   16-NOV-2001; 2001WO-US42950.
XX
PR   17-NOV-2000; 2000US-0714936.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Goodrich RW,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA;
PI   Ren F,  Xue AJ,  Yang Y,  Wehrman T,  Drmanac RT;
XX
DR   WPI; 2002-590824/63.
DR   N-PSDB; ABP64685.
XX
PT   New isolated polynucleotide, useful in research, diagnostic or
PT   therapeutic methods, e.g. preventing or treating disorders involving
PT   aberrant protein expression or biological activity -
XX
PS   Claim 1; SEQ ID 4; 394pp; English.
XX
CC   The present invention relates to novel human coding sequences
CC   (ABQ99268-ABQ99608) and proteins (ABP64682-ABP65022). The sequences are
CC   useful in therapeutic, diagnostic and research methods. The
CC   polynucleotides may be used in the field of molecular biology as
CC   hybridisation probes, primers for PCR, for chromosome and gene mapping,
CC   for the recombinant production of protein, or in generation of anti-sense
CC   DNA or RNA. The polynucleotides are useful in diagnostics as expressed
CC   sequence tags (ESTs) for identifying expressed genes or for physical
CC   mapping of the human genome. The proteins may be used as molecular weight
CC   markers, or as nutritional sources or supplements. The proteins may be
CC   used to maintain and expand cell population in a totipotential or
CC   pluripotential state useful for re-engineering damaged or diseased
CC   tissues, transplantation, manufacture of bio-pharmaceuticals or the
CC   development of bio-sensors. The polynucleotides and proteins are useful
CC   for preventing, treating or ameliorating disorders involving aberrant
CC   protein expression or biological activity, e.g. haematopoietic disorders,
CC   central/peripheral nervous system diseases, mechanical and traumatic
CC   disorders, non-healing wounds, immune deficiencies and disorders,
CC   infectious diseases caused by viral, bacterial or fungal infection,
CC   autoimmune disorders, allergic reactions and conditions, coagulation
CC   disorders, or cancer. The polynucleotide sequences of the invention were
CC   assembled from ESTs isolated mainly by sequencing by hybridisation, and
CC   in some cases, sequences obtained from one or more public databases.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 5469 BP; 1801 A; 983 C; 1097 G; 1588 T; 0 other;

Alignment Scores:
Pred. No.:              2.65e-202     Length:         5469
Score:                  3047.00       Matches:        614
Percent Similarity:     75.53%        Conservative:   133
Best Local Similarity:  62.08%        Mismatches:     179
Query Match:            42.36%        Indels:         64
DB:                     24            Gaps:           16

US-09-512-581B-2 (1-1391) x ABQ99271 (1-5469)

Qy        399 AspLys---ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLys 417
              ||||||   :::|||||||||||||||||||||||||||||||||||||||:::||||||
Db         42 GATAAACTTAAGTGGCGAGTAAGAAAAGAAGCTATGATGGGTCTGGCTCAGCTTTATAAG 101

Qy        418 LysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLys 437
              ||||||   |||            |||||||||:::||||||:::::::::::|||||||||
Db        102 AAATACTGTCTTCATGGTGAAGCAGGAAAGGAAGCTGCAGAGAAAGTCAGCTGGATAAAG 161

Qy        438 AspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArg 457
              ||||||||||||||||||||||||||||||||||||||||||:::||||||||||||:::
Db        162 GACAAACTTCTGCATATTTATTATCAGAACAGCATTGACGACAAACTGTTGGTAGAGAAA 221

Qy        458 IlePheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeu 477
              |||||||||||||||:::|||||||||||||||||||||||||   ||||||||||||||||
Db        222 ATCTTTGCTCAGTATCTTGTCCCCCACAACCTGGAAACAGAAGAGAGAATGAAATGCTTA 281

Qy        478 TyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLys 497
              |||||||||||||||:::|||||||   ||||||||||||||||||||||||||||||||
Db        282 TATTACTTATATGCTAGTTTGGATCCAAATGCTGTAAAAGCTCTCAACGAAATGTGGAAG 341

Qy        498 CysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLys 517
              |||||||||:::||||||      |||::::::|||||||||||||   |||||||||
Db        342 TGTCAGAACATGCTTCGGAGCCATGTACGCGAACTATTGGATTTGCACAAGCAGCCTACA 401

Qy        518 ThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuPro 537
              ::::::|||:::      |||:::|||  |||:::|||   |||    :::|||||||||
Db        402 TCAGAGGCTAACTGTTCTGCCATGTTTGGAAAACTGATGACCATAGCAAAGAATTTGCCT 461

Qy        538 AspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGlu 557
              ||||||||||||||||||||||||:::||||||||||   |||||||||   |||||||||
Db        462 GACCCCGGGAAAGCACAAGATTTTGTGAAGAAATTTAACCAGGTTCTCGGCGATGATGAG 521

Qy        558 LysIleArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGlu 577
              |||:::|||   |||||||||:::|||:::|||||||||||||||||||||||||||:::
Db        522 AAACTTCGGTCTCAGTTGGAGTTATTAATTAGCCCAACCTGTTCTTGCAAACAAGCAGAT 581

Qy        578 GlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPhe 597
                 |||||||||||||||   :::||||||   |||||||||||||||||||||||||||
Db        582 ATTTGTGTGAGAGAAATAGCCCGGAAACTTGCAAATCCTAAGCAACCAACAAATCCTTTT 641

Qy        598 LeuGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSer 617
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||:::
Db        642 CTAGAGATGGTCAAATTTCTGTTGGAAAGAATCGCACCTGTGCACATTGATTCAGAAGCC 701

Qy        618 IleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAsp 637
              ||||||||||||:::|||   :::|||||||||||||:::||||||||||||||||:::
Db        702 ATAAGTGCGCTAGTGAAATTGATGAATAAGTCAATAGAGGGGACAGCAGATGATGAAGAG 761

Qy        638 GluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSer 657
              |||||||||   |||   |||||||||:::||||||||||||||||||||||||||||||
Db        762 GAGGGTGTAAGTCCAGATACAGCTATCCGTTCAGGACTTGAACTTCTTAAGGTTCTGTCT 821

Qy        658 PheThrHisProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeu 677
              ||||||||||||   |||||||||||||||||||||:::||||||||||||   ||||||
Db        822 TTTACACATCCTACCTCGTTCCACTCTGCAGAGACATATGAGTCCTTGTTACAGTGCCTA 881

Qy        678 LysMetAspAspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySer 697
              :::|||:::|||:::|||||||||||||||||||:::||||||||||||:::|||||||||
Db        882 AGAATGGAGGATGACAAGGTAGCAGAAGCTGCTATTCAAATTTTTAGAAATACAGGTCAC 941

Qy        698 LysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLys 717
              |||||||||   |||   |||   |||||||||   |||:::|||:::||||||   |||
Db        942 AAAATAGAAACAGACCTTCCCCAGATACGATCGACCTTAATTCCCATTTTACATCAAAAA 1001

Qy        718 SerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePhe 737
              :::|||:::|||   |||   |||||||||   |||:::|||||||||||||||||||||
Db       1002 GCAAAGAGGGGTACTCCACACCAAGCAAAACAGGCTGTGCACTGTATACACGCCATATTC 1061

Qy        738 SerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspPro 757
              ::::::||||||   |||   ||||||||||||||||||||||||   :::|||||||:::
Db       1062 ACAAATAAAGAAGTCCAGCTTGCACAGATTTTTGAGCCACTCAGTAGGAGTCTGAATGCT 1121

Qy        758 SerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAla 777
                 |||   ||||||||||||||||||||||::::::||||||||||:::::||||||
Db       1122 GATGTGCCAGAACAACTTATAACTCCATTAGTTTCATTGGGCCACATTTCTATGTTAGCA 1181

Qy        778 ProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeu 797
              |||||||||||||||:::|||   ||||||   ||||||   |||||||||||||||||||
Db       1182 CCAGATCAGTTTGCTTCCCCAATGAAATCTGTAGTAGCAAATTTTATTGTGAAAGATCTG 1241

Qy        798 LeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGlu 817
```

```
                 ||||||||||||||      |||:::|||      |||||||||   |||||||||||
Db        1242 CTAATGAATGACAGGTCAACAGGTGAAAAGAATGGAAAACTGTGGTCTCCAGATGAAGAG 1301

Qy         818 ValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeu 837
               |||||||||||  :::   |||:::||||||||||||:::::::||||||||||||||
Db        1302 GTTTCCCCTGAAGTACTAGCAAAGGTACAGGCAATTAAACTTCTGGTAAGGTGGCTGTTG 1361

Qy         838 GlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIle 857
               ||||||||||||||  |||||||||     ||||||||||||||||:::    :::
Db        1362 GGTATGAAAAACAACCAGTCTAAATCTGCCAATTCAACCCTTCGGTTATTATCAGCGATG 1421

Qy         858 LeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArg 877
               |||   |||:::|||||||||||||||||||      :::||||||||   |||||||||||
Db        1422 TTGGTTAGTGAGGGTGACCTGACAGAGCAAAAGAGGATCAGTAAATCTGATATGTCTCGC 1481

Qy         878 LeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGlu 897
               ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||
Db        1482 TTGCGATTAGCTGCTGGTAGTGCCATAATGAAGCTTGCTCAGGAACCTTGTTACCATGAA 1541

Qy         898 IleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGln 917
               |||||||||   ||||||:::||||||||||||||     ||||||||||||||||||||
Db        1542 ATTATTACCCCAGAACAGTTTCAGCTCTGTGCACTTGTTATTAATGATGAGTGTTACCAA 1601

Qy         918 ValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeu 937
               ||||||||:::|||||||||||||||||||||||    |||    :::|||   ||||||||
Db        1602 GTAAGGCAGATATTTGCTCAGAAGCTGCATAAGGCACTTGTGAAGTTACTGCTCCCATTG 1661

Qy         938 GluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHis 957
               |||||||||||||||   |||||||||||||||||||||||||||||||||||||||||
Db        1662 GAGTATATGGCGATCTTTGCCTTGTGTGCCAAAGATCCTGTGAAGGAGAGAAGAGCACAC 1721

Qy         958 AlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAla 977
               |||||||||||||||:::||||||||||:::::|||||||||||||:::||||||:::
Db        1722 GCACGACAATGTTTACTGAAAAATATCAGTATACGCAGGGAATACATTAAGCAGAATCCT 1781

Qy         978 AlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHis 997
                       :::|||||||||||||||||||||||||||||||||||||||||    ||||||
Db        1782 ATGGCTACTGAGAAATTATTATCACTGTTGCCTGAATATGTAGTTCCATACATGATTCAC 1841

Qy         998 LeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspVal 1017
               |||||||||||||||||||||||:::    :::   ||||||::::::||||||:::|||:::
Db        1842 CTGCTAGCCCATGATCCAGATTTTACAAGATCACAAGATGTTGATCAGCTTCGTGATATC 1901

Qy        1018 LysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHis 1037
               |||||||||||||||||||:::||||||:::||||||   |||||||||||||||||||||
Db        1902 AAAGAGTGCCTATGGTTCATGCTTGAAGTTTTAATGACAAAGAATGAAAACAATAGCCAT 1961

Qy        1038 AlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAsp 1057
               ||||||:::::||||||   ||||||||||||   |||:::|||||||||||   ||||||
Db        1962 GCCTTTATGAAGAAGATGGCAGAGAACATCAAGTTAACCAGAGATGCCCAGTCTCCAGAT 2021

Qy        1058 AspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMet 1077
               ::::::|||   |||||||||||||||||||||||||||||||||:::   :::|||
Db        2022 GAATCCAAGACAAATGAAAAACTGTATACAGTATGTGATGTGGCTCTCTGTGTTATAAAT 2081

Qy        1078 SerLysSerThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhe 1097
               |||||||||          :::   :::||||||||||||||||||||||||   :::|||
Db        2082 AGTAAAAGTGCTTTGTGCAATGCAGATTCACCAAAGGACCCAGTCCTCCCAATGAAATTT 2141

Qy        1098 PheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLys 1117
               ||||||||||||:::|||:::|||   |||   |||:::|||:::     |||    :::
Db        2142 TTTACACAACCTGAAAAGGACTTCTGTAACGATAAGAGTTATATTTCAGAAGAGACAAGA 2201

Qy        1118 SerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysPro 1137
                          ||||||||||||        ||||||||||||||||||||||||
Db        2202 GTACTTCTGTTAACAGGAAAGCCAAAGCCTGCTGGAGTACTAGGTGCAGTAAATAAGCCT 2261

Qy        1138 LeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSer--- 1156
               ||||||:::   |||:::::::         :::|||:::      ||||||   |||
Db        2262 TTATCAGCAACGGGAAGGAAACCCTATGTTAGAAGCACTGGCACTGAGACTGGAAGCAAT 2321

Qy        1157 ---AsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArg---Leu 1174
                  |||   :::|||     |||||||||:::        |||   :::   :::
Db        2322 ATTAATGTAAATTCAGAGCTGAACCCTTCAACCGGAAATCGATCAAGGGAACAGAGTTCA 2381

Qy        1175 AspSerSerGluMetAspHisSerGluAsnGluAsp------TyrThrMetSerSerPro 1192
               :::::::::|||          ||||||||||||:::            :::   |||
Db        2382 GAGGCAGCAGAAACTGGAGTTAGTGAAAATGAAGAGAACCCTGTGAGGATTATTTCAGTC 2441

Qy        1193 LeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGlu 1212
                  |||   |||   |||
Db        2442 ACACCTGTAAAGAATATTGAC--------------------------------------- 2462

Qy        1213 LysProArgGlyArgLysLysThrProValThrGluGlnGluGluLysLeuGlyMetAsp 1232
                                       ||||||      :::|||        :::    |||
Db        2463 ------------------------CCAGTAAAGAATAAGGAA------ATTAATTCTGAT 2492

Qy        1233 AspLeuThrLysLeuValGlnGluGlnLysProLysGlySerGlnArgSerArgLysArg 1252
                  |||                  |||              |||   |||   :::||||||
Db        2493 CAGGCTACC---------------CAGGGCAACATCAGCAGTGACCGAGGAAAGAAAAGA 2537

Qy        1253 GlyHisThrAlaSerGluSerAspGlu---GlnGlnTrpProGluGluLysArgLeuLys 1271
                     ||||||:::    ::::::      ||||||        :::||||||
Db        2538 ACAGTAACAGCAGCTGGTGCAGAGAATATCCAACAAAAAACAGATGAGAAA--------- 2588

Qy        1272 GluAspIleLeuGluAsnGluAspGluGlnAsnSerPro---ProLysLysGlyLysArg 1290
                                     ||||||        |||   |||    |||    :::|||
Db        2589 -----------------GTAGATGAATCGGGACCTCCCGCCCCTTCCAAACCCAGGAGA 2630

Qy        1291 GlyArgProProLysProLeuGly---GlyGlyThrProLysGluGluProThrMetLys 1309
               ||||||   ||||||               |||         |||    :::        |||
Db        2631 GGACGTCGACCCAAGTCTGAATCTCAGGGCAATGCTACCAAAAATGATGATCTAAATAAA 2690

Qy        1310 ThrSerLysLysGlySerLysLysLysSer---GlyProProAlaProGluGluGluGlu 1328
                        ||||||   |||:::   :::      |||      :::|||
Db        2691 CCTATTAACAAGGGAAGGAAGAGAGCTGCAGTGGGTCAGGAGAGCCCTGGGGGTTTGGAA 2750

Qy        1329 --------------------------GluGluGluArgGlnSerGlyAsnThrGluGln 1339
                                         ::::::      |||   |||:::|||    :::
Db        2751 GCAGGTAATGCCAAAGCACCCAAACTGCAAGATTT-AGCCAAAAAGGCAGCACCAGCAGA 2809

Qy        1340 Lys------------SerLysSerLysGlnHis------ArgValSerArgArgAlaGln 1353
               |||              :::|||   |||   |||       |||         :::|||:::
Db        2810 AAGACAAATTGACTTACAAAGGTAAAAATGCATTTGCAAAGGGAGAAAATGAAGGCCAAA 2869

Qy        1354 GlnArgAlaGluSerProGluSerSer 1362
               |||:::   ::::::|||   |||:::
Db        2870 CAGAAG---CAGGCTCCAGCTTCTGCA 2893
```

RESULT 8

AAA47423

```
ID   AAA47423 standard; DNA; 3957 BP.
XX
AC   AAA47423;
XX
DT   20-OCT-2000  (first entry)
XX
DE   Sequence encoding human neuron-associated protein.
XX
KW   Neuron associated protein; NEUAP; neurological disorder; epilepsy;
KW   ischemic cerebrovascular disease; stroke; cerebral neoplasm;
KW   Alzheimer's disease; Pick's disease; Huntington's disease;
KW   dementia; Parkinson's disease; demyelinating disease; meningitis;
KW   prion disease; kuru; Creutzfeldt-Jakob disease; neurofibromatosis;
KW   cerebral palsy; muscular dystrophy; central nervous system; CNS;
KW   peripheral nervous system; PNS; myopathy; schizophrenia;
KW   actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;
KW   cirrhosis; hepatitis; mixed connective tissue disease; MCTD;
KW   myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;
KW   autoimmune disease; inflammation; acquired immunodeficiency syndrome;
KW   AIDS; Addison's disease; adult respiratory distress syndrome;
KW   allergy; ankylosing spondylitis; amyloidosis; anaemia; asthma;
KW   Werner syndrome, trauma; human; ds.
XX
OS   Homo sapiens.
XX
```

```
FH   Key             Location/Qualifiers
FT   CDS             140..2704
FT                   /*tag=  a
FT                   /product= Neuron associated protein
XX
PN   WO200034477-A2.
XX
PD   15-JUN-2000.
XX
PF   10-DEC-1999;   99WO-US30408.
XX
PR   11-DEC-1998;   98US-0210083.
PR   11-DEC-1998;   98US-9123456.
PR   09-FEB-1999;   99US-0119365.
PR   16-MAR-1999;   99US-0124687.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Tang YT,  Yue H,  Baughn MR,  Hillman JL,  Lal P,  Au-young J,  Yang J;
PI   Lu DAM,  Azimzai Y;
XX
DR   WPI; 2000-423423/36.
DR   P-PSDB; AAB01382.
XX
PT   New human neuron-associated proteins and polynucleotides encoding them,
PT   useful for diagnosis, treatment and prevention of cell proliferative
PT   disorders including cancer, neuronal and neurological disorders
XX
PS   Claim 9; Page 130-131; 145pp; English.
XX
CC   Human neuron-associated proteins (NEUAP) can be used for for
CC   treating or preventing a disorder associated with decreased
CC   expression or activity of NEUAP. Antagonists of NEUAP are useful for
CC   treating or preventing disorder associated with increased expression
CC   or activity of NEUAP. NEUAP or their fragments or derivatives are
CC   useful for treating neurological disorder such as epilepsy, ischemic
CC   cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's
CC   disease, Pick's disease, Huntington's disease, dementia and
CC   Parkinson's disease. NEUAPs are also useful for treating other
CC   demyelinating diseases, bacterial and viral meningitis, prion
CC   diseases including kuru, Creutzfeldt-Jakob disease, nutritional and
CC   metabolic diseases of the nervous system, neurofibromatosis, other
CC   developmental disorders of the central nervous system, cerebral
CC   palsy, neuroskeletal disorders, autonomic nervous system disorders,
CC   cranial nerve disorders, spinal cord diseases, muscular dystrophy and
CC   other neuromuscular disorders, peripheral nervous system disorders,
CC   inherited, metabolic, endocrine, and toxic myopathies, mental
CC   disorders including mood, anxiety and schizophrenic disorders, a cell
CC   proliferative disorder such as actinic keratosis, arteriosclerosis,
CC   atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective
CC   tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal
CC   haemoglobinuria, cancers of the adrenal gland, bladder, bone,
CC   bone marrow, brain, breast, cervix, and an autoimmune/inflammatory
CC   disorder such as acquired immunodeficiency syndrome (AIDS), Addison's
CC   disease, adult respiratory distress syndrome, allergies, ankylosing
CC   spondylitis, amyloidosis, anemia, asthma, Werner syndrome,
CC   complications of cancer, hemodialysis, and extracorporeal circulation,
CC   viral, bacterial, fungal parasitic, protozoal, and helminthic
CC   infections, and trauma. This sequence was given the Incyte ID no.
CC   2888437CB1.
XX
SQ   Sequence 3957 BP; 1303 A; 723 C; 811 G; 1120 T; 0 other;

Alignment Scores:
Pred. No.:               1.66e-194      Length:          3957
Score:                   2932.50        Matches:         595
Percent Similarity:      74.51%         Conservative:    133
Best Local Similarity:   60.90%         Mismatches:      185
Query Match:             40.77%         Indels:          65
DB:                      21             Gaps:            16

US-09-512-581B-2 (1-1391) x AAA47423 (1-3957)
```

```
Qy      428 AspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsn 447
            :::||||||:::::::::::|||||||||||||||||||||||||||||||||||||||||
Db        2 GAAGCTGCAGAGAAAGTCAGCTGGATAAAGGACAAACTTCTGCATATTTATTATCAGAAC 61

Qy      448 SerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAsn 467
            ||||||||||||:::|||||||||||||||||||||:::|||||||||:::||||||||||
Db       62 AGCATTGACGACAAACTGTTGGTAGAGAAAATCTTTGCTCAGTATCTTGTCCCCCACAAC 121

Qy      468 LeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsn 487
            |||||||||   ||||||||||||||||||||||||||||||||:::||||||   |||
Db      122 CTGGAAACAGAAGAGAGAATGAAATGCTTATATTACTTATATGCTAGTTTGGATCCAAAT 181

Qy      488 AlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLys 507
            ||||||||||||||||||||||||||||||||||||||||:::||||||       |||:::
Db      182 GCTGTAAAAGCTCTCAACGAAATGTGGAAGTGTCAGAACATGCTTCGGATCCATGTACGC 241

Qy      508 AspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSer 527
            :::||||||||||||   ||||||||   ::::::|||:::      |||:::|||
Db      242 GAACTATTGGATTTGCACAAGCAGCCTACATCAGAGGCTAACTGTTCTGCCATGTTTGGA 301

Qy      528 LysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLys 547
            |||:::|||   |||   :::|||||||||||||||||||||||||||||:::|||
Db      302 AAACTGATGACCATAGCAAAGAATTTGCCTGACCCCGGGAAAGCACAAGATTTTGTGAAG 361

Qy      548 LysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuVal 567
            ||||||   |||||||||   ||||||||||||:::|||   |||||||||:::|||:::
Db      362 AAATTTAACCAGGTTCTCGGCGATGATGAGAAACTTCGGTCTCAGTTGGAGTTATTAATT 421

Qy      568 SerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeu 587
            ||||||||||||||||||||||||||||:::   |||||||||||||||   :::||||||
Db      422 AGCCCAACCTGTTCTTGCAAACAAGCAGATATTTGTGTGAGAGAAATAGCCCGGAAACTT 481

Qy      588 GlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGluArg 607
               ||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db      482 GCAAATCCTAAGCAACCAACAAATCCTTTTCTAGAGATGGTCAAATTTCTGTTGGAAAGA 541

Qy      608 IleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsnLys 627
            ||||||||||||||||||||:::|||:::|||||||||||||||:::|||   :::||||||
Db      542 ATCGCACCTGTGCACATTGATTCAGAAGCCATAAGTGCACTAGTGAAATTGATGAATAAG 601

Qy      628 SerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArg 647
            ||||||:::||||||||||||||||||:::||||||||||      |||   |||||||||
Db      602 TCAATAGAGGGGACAGCAGATGATGAAGAGGAGGGTGTAAGTCCAGATACAGCTATCCGT 661

Qy      648 AlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAla 667
            :::||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db      662 TCAGGACTTGAACTTCTTAAGGTTCTGTCTTTTACACATCCTACCTCGTTCCACTCTGCA 721

Qy      668 GluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGluAla 687
            ||||||:::|||||||||||||   |||||||:::|||:::|||:::|||||||||||||||
Db      722 GAGACATATGAGTCCTTGTTACAGTGCCTAAGAATGGAGGATGACAAGGTAGCAGAAGCT 781

Qy      688 AlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHisIleArg 707
            |||:::|||||||||:::|||||||||   |||||||||   |||   |||   ||||||
Db      782 GCTATTCAAATTTTTAGAAATACAGGTCACAAAATAGAAACAGACCTTCCCCAGATACGA 841

Qy      708 SerAlaLeuLeuProValLeuHisHisLysSerLysLysGlyProProArgGlnAlaLys 727
            |||   |||:::|||:::||||||   |||:::|||:::|||   |||   |||||||||
Db      842 TCGACCTTAATTCCCATTTTACATCAAAAAGCAAAGAGGGGTACTCCACACCAAGCAAAA 901

Qy      728 TyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGlnIle 747
               |||:::||||||||||||||||||||:::::::||||||   |||   |||||||||
Db      902 CAGGCTGTGCACTGTATACACGCCATATTCACAAATAAAGAAGTCCAGCTTGCACAGATT 961

Qy      748 PheGluProLeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLeu 767
            ||||||||||||   ::::||||||:::          |||   |||||||||||||||
Db      962 TTTGAGCCACTCAGTAGGAGTCTGAATGCTGATGTGCCAGAACAACTTATAACTCCATTA 1021

Qy      768 ValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSer 787
            |||::::::||||||||||||::::::||||||||||||||||||:::|||   ||||||
Db     1022 GTTTCATTGGGCCACATTTCTATGTTAGCACCAGATCAGTTTGCTTCCCCAATGAAATCT 1081

Qy      788 TrpValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLys 807
```

```
                 ||||||  ||||||||||||||||||||||||||||||||||       |||:::|||
Db      1082 GTAGTAGCAAATTTTATTGTGAAAGATCTGCTAATGAATGACAGGTCAACAGGTGAAAAG 1141

Qy       808 ThrThrLysLeuTrpValProAspGluGluValSerProGluThrMetValLysIleGln 827
                ||||||||  ||||||||||||||||||||||    :::   |||:::|||
Db      1142 AATGGAAAACTGTGGTCTCCAGATGAAGAGGTTTCCCCTGAAGTACTAGCAAAGGTACAG 1201

Qy       828 AlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGly 847
             ||||||||::::::|||||||||||||||||||||||||||||||   |||||||||
Db      1202 GCAATTAAACTTCTGGTAAGGTGGCTGTTGGGTATGAAAAACAACCAGTCTAAATCTGCC 1261

Qy       848 ThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGln 867
             ||||||||||||||||||:::    :::|||  |||:::||||||||||||||||||
Db      1262 AATTCAACCCTTCGGTTATTATCAGCGATGTTGGTTAGTGAGGGTGACCTGACAGAGCAA 1321

Qy       868 GlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleVal 887
                :::|||||||||  |||||||||||||||||||||||||||||||||||||||:::
Db      1322 AAGAGGATCAGTAAATCTGATATGTCTCGCTTGCGATTAGCTGCTGGTAGTGCCATAATG 1381

Qy       888 LysLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGluGlnTyrGlnLeuCys 907
             |||||||||||||||||||||||||||||||||||||||||    ||||||:::|||||||||
Db      1382 AAGCTTGCTCAGGAACCTTGTTACCATGAAATTATTACCCCAGAACAGTTTCAGCTCTGT 1441

Qy       908 AlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHis 927
             ||||||  ||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db      1442 GCACTTGTTATTAATGATGAGTGTTACCAAGTAAGGCAGATATTTGCTCAGAAGCTGCAT 1501

Qy       928 LysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAla 947
             |||    |||    :::|||  ||||||||||||||||||||||||||||||||||||||
Db      1502 AAGGCACTTGTGAAGTTACTGCTCCCATTGGAGTATATGGCGATCTTTGCCTTGTGTGCC 1561

Qy       948 LysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsn 967
             ||||||||||||||||||||||||||||||||||||||||||||:::||||||||||:::
Db      1562 AAAGATCCTGTGAAGGAGAGAAGAGCACACGCACGACAATGTTTACTGAAAAATATCAGT 1621

Qy       968 ValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeu 987
             :::||||||||||||:::||||||:::             :::|||||||||||||||||||
Db      1622 ATACGCAGGGAATACATTAAGCAGAATCCTATGGCTACTGAGAAATTATTATCACTGTTG 1681

Qy       988 ProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLys 1007
             ||||||||||||||||||||||  ||||||||||||||||||||||||||||:::    :::
Db      1682 CCTGAATATGTAGTTCCATACATGATTCACCTGCTAGCCCATGATCCAGATTTTACAAGA 1741

Qy      1008 ValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIle 1027
             ||||||:::::||||||||:::||||||||||||||||||||||||:::||||||:::
Db      1742 TCACAAGATGTTGATCAGCTTCGTGATATCAAAGAGTGCCTATGGTTCATGCTTGAAGTT 1801

Qy      1028 LeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIle 1047
             ||||||  ||||||||||||||||||||||||||:::::::||||||    |||||||||
Db      1802 TTAATGACAAAGAATGAAAACAATAGCCATGCCTTTATGAAGAAGATGGCAGAGAACATC 1861

Qy      1048 LysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThr 1067
             |||   |||:::|||||||||   ||||||::::::|||   |||||||||||||||||
Db      1862 AAGTTAACCAGAGATGCCCAGTCTCCAGATGAATCCAAGACAAATGAAAAACTGTATACA 1921

Qy      1068 ValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSer 1087
             |||||||||||||||:::   :::|||  |||||||||          :::    :::|||
Db      1922 GTATGTGATGTGGCTCTCTGTGTTATAAATAGTAAAAGTGCTTTGTGCAATGCAGATTCA 1981

Qy      1088 ProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsn 1107
             ||||||||||||||||||||         ::|||||||||||||:::|||:::|||  |||
Db      1982 CCAAAGGACCCAGTCCTCCCAATGAAATTTTTTACACAACCTGAAAAGGACTTCTGTAAC 2041

Qy      1108 ThrLysAsnTyrLeuProProGluMetLysSerPhePheThrProGlyLysProLysThr 1127
                |||:::|||:::       |||   :::             ||||||||||||
Db      2042 GATAAGAGTTATATTTCAGAAGAGACAAGAGTACTTCTGTTAACAGGAAAGCCAAAGCCT 2101

Qy      1128 ThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThr 1147
                ||||||||||||||||||||||||||||||:::   |||::::::
Db      2102 GCTGGAGTACTAGGTGCAGTAAATAAGCCTTTATCAGCAACGGGAAGGAAACCCTATGTT 2161

Qy      1148 LysSerSerArgMetGluThrValSer------AsnAlaSerSerSerSerAsnProSer 1165
             :::|||:::       ||||||   |||       |||   :::|||      |||||||||
Db      2162 AGAAGCACTGGCACTGAGACTGGAAGCAATATTAATGTAAATTCAGAGCTGAACCCTTCA 2221

Qy      1166 SerProGlyArgIleLysGlyArg---LeuAspSerSerGluMetAspHisSerGluAsn 1184
             :::      |||   :::   :::      :::::::::|||      |||||||||
Db      2222 ACCGGAAATCGATCAAGGGAACAGAGTTCAGAGGCAGCAGAAACTGGAGTTAGTGAAAAT 2281

Qy      1185 GluAsp------TyrThrMetSerSerProLeuProGlyLysLysSerAspLysArgAsp 1202
             |||:::                :::   |||     |||   |||     |||
Db      2282 GAAGAGAACCCTGTGAGGATTATTTCAGTCACACCTGTAAAGAATATTGAC--------- 2332

Qy      1203 AspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThrProVal 1222
                                                                   ||||||
Db      2333 ------------------------------------------------------CCAGTA 2338

Qy      1223 ThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGluGlnLys 1242
                :::|||            |||   :::                    |||
Db      2339 AAGAATAAGGAA------ATTAATTCTGATCAGGCTACC--------------CAGGGC 2377

Qy      1243 ProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGlu--- 1261
                   |||   |||   :::||||||    ||||||:::    ::::::
Db      2378 AACATCAGCAGTGACCGAGGAAAGAAAAGAACAGTAACAGCAGCTGGTGCAGAGAATATC 2437

Qy      1262 GlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGln 1281
             ||||||      :::||||||                         ||||||
Db      2438 CAACAAAAAACAGATGAGAAA-------------------------GTAGATGAATCG 2470

Qy      1282 AsnSerPro---ProLysLysGlyLysArgGlyArgProProLysProLeuGly---Gly 1299
                   |||   |||   |||   :::|||||||||   ||||||         |||
Db      2471 GGACCTCCCGCCCCTTCCAAACCCAGGAGAGGACGTCGACCCAAGTCTGAATCTCAGGGC 2530

Qy      1300 GlyThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLysSer 1319
                   |||         |||            ||||||   |||:::   :::
Db      2531 AATGCTACCAAAAAATGATGATCTAAATAAACCTATTAACAAGGGAAGGAAGAGAGCTGCA 2590

Qy      1320 ---GlyProProAlaProGluGluGluGlu-------------------------Glu 1329
                |||      :::|||      |||                              :::
Db      2591 GTGGGTCAGGAGAGCCCTGGGGGTTTGGAAGCAGGTAATGCCAAAGCACCCAAACTGCAA 2650

Qy      1330 GluGluArgGlnSerGlyAsnThrGluGlnLys-----------SerLysSerLysGln 1345
             :::      |||   |||:::|||   :::|||             :::|||   |||
Db      2651 GATTT-AGCCAAAAAGGCAGCACCAGCAGAAAGACAAATTGACTTACAAAGGTAAAAATG 2709

Qy      1346 His------ArgValSerArgArgAlaGlnGlnArgAlaGluSerProGluSerSerAla 1363
             |||      |||         :::|||:::|||:::   :::::::|||   |||:::
Db      2710 CATTTGCAAAGGGAGAAAATGAAGGCCAAACAGAAG---CAGGCTCCAGCTTCTGCAAAA 2766

Qy      1364 ------IleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgPro 1378
                   |||               |||      :::   :::   ||||||
Db      2767 ACTTGGATTCACAAATGTCCCTGAACAGAAAATGAAGCTCAACGCAGACCG 2817
```

RESULT 9
ABL22105

```
ID   ABL22105 standard; DNA; 3920 BP.
XX
AC   ABL22105;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 17788.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
```

```
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 17788; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3920 BP; 1053 A; 1007 C; 947 G; 913 T; 0 other;

Alignment Scores:
Pred. No.:              1.42e-128     Length:        3920
Score:                  1986.50       Matches:       448
Percent Similarity:     57.77%        Conservative:  258
Best Local Similarity:  36.66%        Mismatches:    441
Query Match:            27.62%        Indels:        76
DB:                     23            Gaps:          22

US-09-512-581B-2 (1-1391) x ABL22105 (1-3920)

Qy      13 IleThrTyrProProGlyValLysGluIleSerAspLysIleSerLysGluGluMetVal 32
Db     273 ATAGTGTACCCCACCGGATGCCGGCCGCTGGTCGAGGATTTGGGCACAGACGAGCTGATA 332

Qy      33 ArgArgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSerGluGluGlu 52
Db     333 AGGCGCCTCAAGACCCTCGCCAATGTTCTGCAGACCATGGACCAGGACGACAATCTCTAC 392

Qy      53 LysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPhePheLeuLysHisPro 72
Db     393 CAGCAA---TACATACCATTGGCCCTCCATCTGCTGGACGACTTCTTTATGCAACATCCA 449

Qy      73 GlyLysAspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAla 92
Db     450 TCCCGTGATGTCCAGCTCTTAATCGCGTGCTGCGTGGCGGATGTCCTGCGGGTTTATGCG 509

Qy      93 ProGluAlaProTyrThrSerProAspLysLeuLysAspIlePheMetPheIleThrArg 112
Db     510 CCAGAGGCTCCCTACAAAGAGCAAGATCAAATCAAAACTATATTTAAGTTTTTATCAAG 569

Qy     113 GlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeu 132
Db     570 CAACTGCACGGTCTGAAGGATCCGCGCGATCCCTCTTTCAAGCGTTACTTCTATCTGCTG 629

Qy     133 GluAsnIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSerAsnGlu 152
Db     630 GAGAACCTTGCTTTTGTCAAGTCTTTTAATATGTGTTTCGAACTGGAGGACTGCCAAGAA 689

Qy     153 IlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAsnGlyHisAsnGlnLys 172
Db     690 ATCTTCCAAGACCTGTTTAGTACCATCTTCAAGATTGTCAATGACCAGCACAGTGTCAAG 749

Qy     173 ValHisMetHisMetValAspLeuMetSerSerIleIleCysGluGlyAspThrValSer 192
Db     750 GTGACTAACTTTTTCTAGATGTGCTAAGCCCGTTGATCACAGAAGCCGACAATCTGTCG 809

Qy     193 GlnGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAsnLys 212
Db     810 GTGGAATTGTTGGATCTCATTCTGATTAACATCGTAGAGCCGTACAAATCTAACAACAAG 869

Qy     213 GlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTyr 232
Db     870 TTTGCCTGCCAATTAACAGAACAACTTCTCACTAAAACGGGCGATGCTCTCGAGTCCACC 929

Qy     233 IleThrThrPhePheAsnGlnValLeuMetLeuGlyLys------ThrSerIleSerAsp 250
Db     930 ATCAAAATGTTCTTTAACCGTGCTCTGGTCATGGATAAGCCAAACACAAAACTGTCC--- 986

Qy     251 LeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeu 270
Db     987 ATTACAAACAAAATCTATGATATCATTTACGAGCTTAACCGCATCAATGCAGGCTTGTTG 1046

Qy     271 LeuSerValLeuProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluGluArgLeu 290
Db    1047 TGTTCAGTGCTACCTCAGCTGGAGAACAAACTGCTGTCCACGGACGATGCTGAGAGGCTA 1106

Qy     291 GlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerGln 310
Db    1107 AAGGCAACCACTCTCTTGTCTCGTATGTTCTCCGAAAAGGACTCGCAACTAGCGAAAAAG 1166

Qy     311 AsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIleArg 330
Db    1167 TACCCCAACCTGTTGAAAATCTTCTTTGGGCGCTTCTGCGACATCACCGAACCAGTTCGC 1226

Qy     331 LeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAsp 350
Db    1227 ATCAAGTGTGTTCAGTCATCCATGCACTTCCTGCTCAATCACCCAAGTCTTCAGCACGAT 1286

Qy     351 LeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleArgHisAspVal 370
Db    1287 ATCACCGAGAAATTGCGGCTTCGAAATCACGATCTTGACGAAGTGGTCCGCCACGAGGTG 1346

Qy     371 IleValSerIleValThrAlaAlaLysLysAspIleLeuLeuValAsnAsp------His 388
Db    1347 GTAATGGCTATTGTGGAAACTGCCAAGCGCGACTTCACCCTCGTTCTCGAAGCGCCCGAT 1406

Qy     389 LeuLeuAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLysGluAla 408
Db    1407 CTACTAGAAATTGTGCGCGAGCGGACGCTAGATAAGAAGTACAAAATTCGCAGGGATGCG 1466

Qy     409 MetMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAsp 428
Db    1467 ATGAATGGTCTGGCCTACATCTACAAGCGCGCAATTTGCGAACCCAACGATCTAAGCACT 1526

Qy     429 AlaAlaLys---GlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsn 447
Db    1527 GGCCTTAAGGTCAGGGTTGACTGGATTAAGAACAAGATACTACATGGATACTACAAAGTG 1586

Qy     448 SerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAsn 467
Db    1587 GGCTTGGAGGACCGCCTGCTTGTGGAACGCCTACTTATCACCTGCCTGGTTCCCTATAAA 1646

Qy     468 LeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsn 487
Db    1647 CTAGCTCCGGAAGAGCGCATGAAGAAGCTATATCATTTGCTAGGCGATCTCGATGCCAAT 1706

Qy     488 AlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLys 507
Db    1707 GCCACCAAGGCTTTCGTCGAGCTGCAGAAGAACCAAATGAAGACACGCAATACGGTCAGT 1766

Qy     508 AspLeuLeuAspLeuIleLysGlnProLys---ThrAspAlaSerValLysAlaIlePhe 526
Db    1767 GATTGGATCAAGCTACACCACTCTAAGGAGTTCACGCCGCGCGTACTTAGTCAGCTCAGC 1826

Qy     527 SerLysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMet 546
Db    1827 GCCAAGCAGGCCAACATTGCCAAACTGCTTCCAGATCCCCTAAAAGCGGCAGAGTACCTA 1886

Qy     547 LysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeu 566
```

```
                 :::|||:::          |||        |||    ::::::    :::    :::    ::::::
Db       1887 ACCCAGTTTAGCAACAACCTGCGAAAAGATGCTCAGCTCCTACGATGCATTAACATTGTC 1946

Qy        567 ValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLys 586
              :::                 ||||||::::::              :::      :::    ||||||
Db       1947 CTTAAACGCGACGTAAGCTGTCGGGAGTGTGCTGACACGATGGAGTTCTCCTGAAAAAG 2006

Qy        587 LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGlu 606
              ||||||          |||    :::|||    :::              :::|||   |||:::|||
Db       2007 CTTGGCGCCCACGTCCAA---TCGAATCTGTATTACAACACAGTTAAGATGCTGATTGAG 2063

Qy        607 ArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsn 626
              |||:::|||    |||    :::|||    ||||||||    ||||||         :::
Db       2064 CGCGTGGCATCAGTGATGGTGGACAAGGAGTCTATTGGCGTACTTATTAGCCTAATTGAA 2123

Qy        627 LysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIle 646
              :::   |||:::          :::         |||:::   |||:::          :::|||
Db       2124 CAATGCATCGAGAAAGGGTCCATGTGTGAAGAGATCGGAATTTCGGCTCAGGAAGCGGGC 2183

Qy        647 ArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSer 666
                    ||||||:::||||||    :::||||||:::                   |||    :::
Db       2184 GAGCGCGGACTCAAGTTGTTGGCCATGCTTTCATACGTCTTTTCGGCTCATTTCTTCACC 2243

Qy        667 AlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGlu 686
                    :::         |||::::::    |||         :::    :::    ||||||
Db       2244 GATACTTCGCTGCGTCATTTAATTTCCCTGCTCAGTTACGAGCAGGATTATGTTGCACCA 2303

Qy        687 AlaAlaLeuGlnIlePheLysAsnThrGly---------SerLysIleGluGluAspPhe 703
                       |||:::          :::    |||                |||::::::
Db       2304 CTGGTGCTAAAGACACTAACTCATCTGGGGCGCTATCAGCCACTCATTGACGACCCTACA 2363

Qy        704 ProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLysGlyProPro 723
              |||    |||              |||    ||||||         :::       |||    |||
Db       2364 CCGGCTATTCTCGACGAGCTGGCTCCGGTGTGCAAGGACTTCGCACTTATTGGAACCCCC 2423

Qy        724 ArgGlnAlaLysTyrAlaIleHisCysIleHis------------------------- 734
              :::|||||||||:::|||:::    ||||||
Db       2424 AAGCAAGCAAAGCACGCGGTGCGATGTATTTCGTAAACAGCCAGTCGTCGGCTTCCACT 2483

Qy        735 -----------------AlaIlePheSerSerLysGluThrGln-------------- 743
                                |||       :::|||       ||||||
Db       2484 GATGGAGCAACAAGTGGAGCTGGAAGTGCGTCCACAACTACACAAACAGTGCATCCTATT 2543

Qy        744 PheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeuGluHisLeu 763
              |||    :::|||   |||    |||             |||    |||    |||    ||||||
Db       2544 TTCAACGAGATCATCGAGACGCTTCGCCTAAAACTGACACCA---AACTGTGAGCATCAG 2600

Qy        764 IleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAla 783
                 |||    :::||||||:::||||||||||||           |||       |||
Db       2601 CGCACAAAGATTGTGACCTTGGGTCACATTGCCTACAATATGCCACAGGCCTTCCTAACG 2660

Qy        784 ProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeu 803
              |||    |||:::    :::|||      ||||||||:::||||||:::    :::    :::
Db       2661 CCCATTAAGAATATGATTGCGCGACGAATTGTCAAAGAGCTGCTTATCCAAGAA---GTT 2717

Qy        804 ProGlyLysLysThrThrLysLeu--------------TrpValProAspGluGluVal 818
              |||    ::::::       :::|||                |||          |||::::::
Db       2718 CCTGCGCAGCGAGACTACGAACTGCCAGAGGACAGTGATTGGTGTGCCCAAGAGAAACTA 2777

Qy        819 SerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGly 838
                 |||:::|||:::    |||:::    |||:::|||    |||    |||||||||||||||
Db       2778 CCGCCGGACACTCTATGTAAGCTGGATGCGCTCAAGGCTATGGCCAGGTGGCTTTTGGGA 2837

Qy        839 MetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeu 858
              ::::::    :::          :::       |||    |||:::|||             :::
Db       2838 TTGCGTACCGAT---GAGCACGCTGCCCAGAAAACATTCCGAATGCTGGCGGCGTTTGTT 2894

Qy        859 HisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeu 878
              :::      ||||||||      |||    ::::::          :::    |||    |||
Db       2895 AACCAACGAGGCGATTTGCTCGGCCAGAACCGTCTTTGCGGTGCCGAGAAATCTTGGCTG 2954

Qy        879 ArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIle 898
              ||||||    |||       |||::::::|||:::    ::::::                :::
Db       2955 CGCCTCGGGGCAGCCTGCGCCATGCTCAAAGTGTGCGAACAAAAGGGCGTAGGTGATCAG 3014

Qy        899 IleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnVal 918
                 :::   |||||||||       :::       :::    |||          :::|||
Db       3015 TACAGCGCTGAGCAGTATTTGCAGCTTTCCCAGCTGATGGCTGATCCGGTGCCAGAAGTT 3074

Qy        919 ArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArg----------- 934
              |||::::::||||||:::|||||||||||||||||||||
Db       3075 CGGGAAATCTTTGCTCGCAAGCTGCACAAAGGATTAAGCAGAAGTTTGCCCAGGAACTGT 3134

Qy        935 LeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArg 954
              |||||||||::::::|||    :::          |||          :::        ::::::
Db       3135 TTGCCGCTGGACTTCATGGGCTTGTATGTGCTGGCTGGTCTAGAGACTGAGAGGAAATTG 3194

Qy        955 ArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLys 974
              :::                  |||            ::::::|||   ||||||||||||||||
Db       3195 CAAGACCTTGTGCGTCACTATGCAGAAACGGATGTAAACAAACGGCGGGAATATCTCAAG 3254

Qy        975 GlnHisAlaAlaValSer-----------------GluLysLeuLeuSerLeuLeuPro 988
                 |||    |||                         :::    |||    :::||||||
Db       3255 ACTGTGGCTATGACATCTCCCGACAGCTCAACGGAGTCACAATCATTACACATACTACCT 3314

Qy        989 GluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysVal 1008
              :::|||::::::    :::    |||    :::|||    |||||||||       :::
Db       3315 GACTACATGCTGGCTTTCGCTATTCCCGTGCTGGTCCACGATCCACGCTTCACGAATCAC 3374

Qy       1009 GlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeu 1028
              :::|||       ||||||:::    :::::::::||||||    |||:::||||||    |||
Db       3375 GAGGACTACGTACAGCTGCGCAAGATGGAGAAGTGCCTGCGTTTCATTCTGGAGCCGCTG 3434

Qy       1029 MetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLys 1048
              |||||||||    |||          |||:::|||    :::::::::::::        ||||||
Db       3435 ATGGCCAAACGAGAAACGTTTGTCCATAGCTTCTACAAGCAGCTGCTGCAGCTGATAAAG 3494

Qy       1049 GlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrVal 1068
                 :::    :::    |||        |||    :::    |||    |||::::::    :::
Db       3495 CATCGCGAGTTCAGTCTGGGG---TCGGACAAGCGCGACAACTATAAAATGTGGGCGCTC 3551

Qy       1069 CysAspValAlaMetAsnIleIleMetSerLys---------------------SerThr 1081
              ||||||:::||||||    ||||||    ||||||                   ::::::
Db       3552 TGCGATCTTGCCATGTACATTATCGACTCCAAGTTCAGTCCATTTGATGGCAACACGAGC 3611

Qy       1082 ThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGlnPro 1101
              |||:::|||:::                          ||||||       ::::::    :::|||
Db       3612 ACCTTTTCCATGCCGCTG-----------GCTTTGCCAGAAATGTATTATAAAGAGCCT 3659

Qy       1102 Asp---LysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe 1120
                    ||||||    |||          |||:::|||    ::::::                :::
Db       3660 GCCGTTGCGAATTTCCAAAACAATGACGTCTATATACCGCTGGACGTG---------TAT 3710

Qy       1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
              |||    |||          |||:::                   |||       :::::::|||
Db       3711 ACGCTGGGAGCCAAATCCACGAGCAAAGCTGCCGCAACAGC-AATGACAACGTCGCGAGC 3769

Qy       1141 AlaGlyLysGlnSerGlnThr--------------------------LysSerSerArg 1151
              :::             :::|||                         ||||||:::
Db       3770 AGCAGTGGCTCCAAAGAGACCGGCGGAACAGTCAATCATGGATGATGAAAATCCGCAGGA 3829

Qy       1152 MetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLys 1171
                 ::::::|||          :::|||       :::       :::    |||:::
Db       3830 GAACAATCTGTTCGACAACATACGAGCGGCTGACACTACGGAGCCCATGGCCAAACGAAC 3889

Qy       1172 GlyArg 1173
                 |||
Db       3890 GCGCGC 3895
```

RESULT 10
ABL22104/c
ID   ABL22104 standard; DNA; 6990 BP.
XX
AC   ABL22104;
XX
DT   26-MAR-2002 (first entry)

```
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 17785.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 17785; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 6990 BP; 1907 A; 1513 C; 1586 G; 1984 T; 0 other;

Alignment Scores:
Pred. No.:              2.48e-110     Length:        6990
Score:                  1729.50       Matches:       463
Percent Similarity:     49.43%        Conservative:  270
Best Local Similarity:  31.22%        Mismatches:    436
Query Match:            24.04%        Indels:        318
DB:                     23            Gaps:          37

US-09-512-581B-2 (1-1391) x ABL22104 (1-6990)

Qy         13 IleThrTyrProProGlyValLysGluIleSerAspLysIleSerLysGluGluMetVal 32
              |||   ||||||  |||   :::   :::   :::   :::      :::|||::::::
Db       5398 ATAGTGTACCCCACCGGATGCCGGCCGCTGGTCGAGGATTTGGGCACAGACGAGCTGATA 5339

Qy         33 ArgArgLeuLysMet-ValVal------------------------------------- 39
              ||||||||||||::: |||:::
Db       5338 AGGCGCCTCAAGGTGAGTCCTGCCCCTGGTTCCCTTAGCAAAACTTCTCACTTAGAATGA 5279

Qy         40 ---------------------LysThrPheMetAspMetAspGlnAspSerGluGluGl 52
                                                   ||||||||||||
Db       5278 TTCTGTTTTACAGACCCTCGCCAATGTTCTGCAGACCATGGACCAGGACGACAATCTCTA 5219

Qy         52 uLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPhePheLeuLysHisPr 72
                ::::::   |||:::   |||||||||||||||         |||||||||:::::|||||
Db       5218 CCAGCAA---TACATACCATTGGCCCTCCATCTGCTGGACGACTTCTTTATGCAACATCC 5162

Qy         72 oGlyLysAspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAl 92
              |    :::||||||:::||||||:::|||||||||:::||||||:::     |||:::|||||
Db       5161 ATCCCGTGATGTCCAGCTCTTAATCGCGTGCTGCGTGGCGGATGTCCTGCGGGTTTATGC 5102

Qy         92 aProGluAlaProTyrThrSerProAspLysLeuLysAspIlePheMetPheIleThrAr 112
              |||||||||||||||        |||:::::|||   ||||||   |||    ::
Db       5101 GCCAGAGGCTCCCTACAAAGAGCAAGATCAAATCAAAACTATATTTAAGTTTTTATCAA 5042

Qy        112 gGlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLe 132
              :||||||   ||||||:::|||   :::   |||   |||   ||||||||||||||||
Db       5041 GCAACTGCACGGTCTGAAGGATCCGCGCGATCCCTCTTTCAAGCGTTACTTCTATCTGCT 4982

Qy        132 uGluAsnIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSerAsnGl 152
              ||||||:::|||:::|||||||||:::|||:::|||||||||||||||||         ||
Db       4981 GGAGAACCTTGCTTTTGTCAAGTCTTTTAATATGTGTTTCGAACTGGAGGACTGCCAAGA 4922

Qy        152 uIlePheThrGlnLeuTyrArgThrLeuPheSerValIle------------------- 165
              ||||||    |||:::   |||:::|||   ::::::
Db       4921 AATCTTCCAAGACCTGTTTAGTACCATCTTCAAGATTGTCAAGTAAGTCGAGCTCTTCCT 4862

Qy        166 ------------------------------------------AsnAsnGlyHisAsnGl 171
                                                        ::::::   |||:::
Db       4861 TTATCCTGCGAACTCTTTATAATCACATGTTATATAATTTTCTAGTGACCAGCACAGTGT 4802

Qy        171 nLysValHisMetHisMetValAspLeuMetSerSerIleIleCysGluGlyAspThrVa 191
               ||||||       :::|||:::::|||   :::|||   |||   |||   ::
Db       4801 CAAGGTGACTAACTTTTTCTAGATGTGCTAAGCCCGTTGATCACAGAAGCCGACAATCT 4742

Qy        191 lSerGlnGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAs 211
              :|||   ||||||||||||   :::|||:::|||:::|||      :::|||:::   ||
Db       4741 GTCGGTGGAATTGTTGGATCTCATTCTGATTAACATCGTAGAGCCGTACAAATCTAACAA 4682

Qy        211 nLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGlu-P 231
              ||||   |||      |||   :::   ||||||   :::|||      |||:::||| |
Db       4681 CAAGTTTGCCTGCCAATTAACAGAACAACTTCTCACTAAAACGGGCGATGCTCTCGAGTC 4622

Qy        231 ro-----------------------------TyrIleThrThr---------------- 235
              ||                             :::  ||||||
Db       4621 CACCATCAAAATGGTAATATTTCGTTGATCTTTGGTGCACCACGTTATTCAACCTAATCC 4562

Qy        236 -----------PhePheAsnGlnValLeuMetLeuGlyLys------ThrSerIleSerA 250
                         |||||||||:::   |||::::::   |||      |||   :::|||
Db       4561 CTGATCTACAGTTCTTTAACCGTGCTCTGGTCATGGATAAGCCAAACACAAAACTGTCC- 4503

Qy        250 spLeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuL 270
                ::::::      ::::::|||:::|||   ||||||      |||::::::   ||||
Db       4502 --ATTACAAACAAAATCTATGATATCATTTACGAGCTTAACCGCATCAATGCAGGCTTGT 4445

Qy        270 euLeuSerValLeuProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluGluArgL 290
              ||   |||||||||||||||||||||   ||||||   |||   |||:::   ||||||
Db       4444 TGTGTTCAGTGCTACCTCAGCTGGAGAACAAACTGCTGTCCACGGACGATGCTGAGAGGC 4385

Qy        290 eu---------------------------------------------------------- 290
              ||
Db       4384 TAAGTGAGTAATTTATCAGCATGTACCTCTATTTGCATATTAATAAACATATTTTCGTTA 4325

Qy        291 -----GlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaS 309
                  :::          ||||||:::::||||||      ||||||||||::::||||||
Db       4324 CTACAGAGGCAACCACTCTCTTGTCTCGTATGTTCTCCGAAAAGGACTCGCAACTAGCGA 4265

Qy        309 erGlnAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProI 329
                :::         |||   :::   :::   ||||||||   ||||||      |||:
Db       4264 AAAAGTACCCCAACCTGTTGAAAATCTTCTTTGGGCGCTTCTGCGACATCACCGAACCAG 4205

Qy        329 leArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaL 349
              ::|||::::::|||||||:::   :::   |||   |||:::||||||||||   |||
Db       4204 TTCGCATCAAGTGTGTTCAGTCATCCATGCACTTCCTGCTCAATCACCCAAGTCTTCAGC 4145

Qy        349 ysAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleArgHisA 369
                |||:::||||||   |||::::::|||:::||||||   :::|||   :::|||||:
Db       4144 ACGATATCACCGAGAAATTGCGGCTTCGAAATCACGATCTTGACGAAGTGGTCCGCCACG 4085

Qy        369 spValIleValSerIleValThrAlaAlaLysLysAspIleLeuLeuValAsnAsp---- 387
              ::|||:::::::::||||||                 ||||||     :::
Db       4084 AGGTGGTAATGGCTATTGTGGAAACTGCCAAGCGCGACTTCACCCTCGTTCTCGAAGCGC 4025

Qy        388 --HisLeuLeuAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLysG 407
```

```
Db   4024 CCGATCTACTAGAAATTGTGCGCGAGCGGACGCTAGATAAGAAGTACAAAATTCGCAGGG 3965

Qy    407 luAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyL 427
Db   3964 ATGCGATGAATGGTCTGGCCTACATCTACAAGCGCGCAATTTGCGAACCCAACGATCTAA 3905

Qy    427 ysAspAlaAlaLys---GlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrG 446
Db   3904 GCACTGGCCTTAAGGTCAGGGTTGACTGGATTAAGAACAAGATACTACATGGATACTACA 3845

Qy    446 lnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProH 466
Db   3844 AAGTGGGCTTGGAGGACCGCCTGCTTGTGGAACGCCTACTTATCACCTGCCTGGTTCCCT 3785

Qy    466 isAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspL 486
Db   3784 ATAAACTAGCTCCGGAAGAGCGCATGAAGAAGCTATATCATTTGCTAGGCGATCTCGATG 3725

Qy    486 euAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnV 506
Db   3724 CCAATGCCACCAAGGCTTTCGTCGAGCTGCAGAAGAACCAAATGAAGACACGCAATACGG 3665

Qy    506 alLysAspLeuLeuAspLeuIleLysGln-----------ProLysThrAspAlaSerV 522
Db   3664 TCAGTGATTGGATCAAGCTACACCACTCTAAGGAGTTCACGCCGCGCGTACTTAGTCAGC 3605

Qy    522 alLysAla--------------------------------------------------- 524
Db   3604 TCAGCGCCAAGCAGGCCAACATTGCCAAGTGGGTTCTATAATAAGTTTCTTTTGATAGT 3545

Qy    525 --IlePheSerLysValMetValIleThr-ArgAsnLeuProAspProGlyLysAlaGln 543
Db   3544 TTGTAACCAGTGATGTGTTTTTTTTCTTCAGACTGCTTCCAGATCCCCTAAAAGCGGCA 3485

Qy    544 AspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeu 563
Db   3484 GAGTACCTAACCCAGTTTAGCAACAACCTGCGAAAAGATGCTCAGCTCCTACGATGCATT 3425

Qy    564 GluValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIle 583
Db   3424 AACATTGTCCTTAAACGCGACGTAAGCTGTCGGGAGTGTGCTGACACGATGGGAGTTCTC 3365

Qy    584 ThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPhe 603
Db   3364 CTGAAAAAGCTTGGCGCCCACGTCCAA---TCGAATCTGTATTACAACACAGTTAAGATG 3308

Qy    604 LeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLys 623
Db   3307 CTGATTGAGCGCGTGGCATCAGTGATGGTGGACAAGGAGTCTATTGGCGTACTTATTAGG 3248

Qy    624 -------------------------------------------------------Gln 624
Db   3247 TAAATTGCTTTTAAGTGCCCTTTTGCCAGACTTAATCCGGTAATCTTTTTGCTCAGCCTA 3188

Qy    625 ValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGln 644
Db   3187 ATTGAACAATGCATCGAGAAAGGGTCCATGTGTGAAGAGATCGGAATTTCGGCTCAGGAA 3128

Qy    645 AlaIleArgAlaGlyLeuGluLeuLeu-------------------------------- 653
Db   3127 GCGGGCGAGCGCGGACTCAAGTTGTTGGCCGTGAGTATAGAGATAATTTAAGTTATGATA 3068

Qy    654 ------------------------------LysValLeuSerPheThrHisProIleSer 663
Db   3067 CAGAAATCCATTTGATTGCTTTATACTTGCAGATG-CTTTCATACGTCTTTTCGGCTCAT 3009

Qy    664 PheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLys 683
Db   3008 TTCTTCACCGATACTTCGCTGCGTCATTTAATTTCCCTGCTCAGTTACGAGCAGGATTAT 2949

Qy    684 ValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGly---------SerLysIleGlu 700

Db   2948 GTTGCACCACTGGTGCTAAAGACACTAACTCATCTGGGGCGCTATCAGCCACTCATTGAC 2889

Qy    701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLys 720
Db   2888 GACCCTACACCGGCTATTCTCGACGAGCTGGCTCCGGTGTGCAAGGACTTCGCACTTATT 2829

Qy    721 GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHis----------------- 734
Db   2828 GGAACCCCCAAGCAAGCAAAGCACGCGGTGCGATGTATTTCGTAAACAGCCAGTCGTCG 2769

Qy    735 --------------------------AlaIlePheSerSerLysGluThrGln------ 743
Db   2768 GCTTCCACTGATGGAGCAACAAGTGGAGCTGGAAGTGCGTCCACAACTACACAAACAGTG 2709

Qy    744 ---------PheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
Db   2708 CATCCTATTTTCAACGAGATCATCGAGACGCTTCGCCTAAAACTGACACCA---AACTGT 2652

Qy    761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
Db   2651 GAGCATCAGCGCACAAAGATTGTGACCTTGGGTCACATTGCCTACAATATGCCACAGGCC 2592

Qy    781 PheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
Db   2591 TTCCTAACGCCCATTAAGAATATGATTGCGCGACGAATTGTCAAAGAGCTGCTTATCCAA 2532

Qy    801 AspArgLeuProGlyLysLysThrThrLysLeu--------------TrpValProAsp 815
Db   2531 GAA---GTTCCTGCGCAGCGAGACTACGAACTGCCAGAGGACAGTGATTGGTGTGCCCAA 2475

Qy    816 GluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrp 835
Db   2474 GAGAAACTACCGCCGGACACTCTATGTAAGCTGGATGCGCTCAAGGCTATGGCCAGGTGG 2415

Qy    836 LeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThr 855
Db   2414 CTTTTGGGATTGCGTACCGAT---GAGCACGCTGCCCAGAAACATTCCGAATGCTGGCG 2358

Qy    856 ThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMet 875
Db   2357 GCGTTTGTTAACCAACGAGGCGATTTGCTCGGCCAGAACCGTCTTTGCGGTGCCGAGAAA 2298

Qy    876 SerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyr 895
Db   2297 TCTTGGCTGCGCCTCGGGGCAGCCTGCGCCATGCTCAAAGTGTGCGAACAAAAGGGCGTA 2238

Qy    896 HisGluIleIleThrLeuGluGlnTyr---GlnLeuCysAlaLeuAlaIleAsnAspGlu 914
Db   2237 GGTGATCAGTACAGCGCTGAGCAGTATTTGCAGCTTTCCCAGCTGATGGTGAGTAGAGAC 2178

Qy    914 ----------------------------------------------------------- 914
Db   2177 ATTCATCTTGTGCTTAGAACTTAGTCTCAATGTCTATCTCGGTGTTCCCGTAGGCTGATC 2118

Qy    915 ---CysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeu 933
Db   2117 CGGTGCCA-GAAGTTCGGGAAATCTTTGCTCGCAAGCTGCACAAAGGATTAAGCAGAAGT 2059

Qy    934 Arg------------LeuProLeuGluTyrMetAlaIleCysAlaLeuCysAla------ 947
Db   2058 TTGCCCAGGAACTGTTTGCCGCTGGACTTCATGGGCTTGTATGTGCTGGCTGGTCTAGAG 1999

Qy    948 ---------LysAspProValLysGluArgArgAlaHisAlaArg-GlnCysLeuValLy 964
Db   1998 ACTGAGAGGAAGTGAGTAGTTCAAGAAACCATCCGATCCGTTCGGAGAAGTATTGATAAA 1939

Qy    964 s------------------------------------------AsnIleAsnValAr 969
Db   1938 TATTTCAACAGATTGCAAGACCTTGTGCGTCACTATGCAGAAACGGATGTAAACAAACG 1879

Qy    969 gArgGluTyrLeuLys------------------------------------------ 974
Db   1878 GCGGGAATATCTCAAGACTGTGGCTATGACATGTAGGTGTTTTACTTTGGAATCATATT 1819
```

```
Qy        975 ---------------------------------GlnHis-AlaAlaValSerGluLysL 983
                                                 ::::::           :::
Db       1818 GTGTATTTTAACCATTAAATCAAATTATTCCAGCTCCCGACAGCTCAACGGAGTCACAAT 1759

Qy        983 euLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspP 1003
                |||   :::||||||:::|||::::::   :::    |||    :::|||   |||||||
Db       1758 CATTACACATACTACCTGACTACATGCTGGCTTTCGCTATTCCCGTGCTGGTCCACGATC 1699

Qy       1003 roAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpP 1023
              ||   :::        :::|||     ||||||:::   :::::::::||||||     |
Db       1698 CACGCTTCACGAATCACGAGGACTACGTACAGCTGCGCAAGATGGAGAAGTGCCTGCGTT 1639

Qy       1023 heValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysM 1043
              ||:::||||||   |||||||||||   |||             |||:::|||   :::::::
Db       1638 TCATTCTGGAGCCGCTGATGGCCAAACGAGAAACGTTTGTCCATAGCTTCTACAAGCAGC 1579

Qy       1043 etValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsn- 1062
              ::::::::    ||||||     :::   :::    |||       |||    :::    |||
Db       1578 TGCTGCAGCTGATAAAGCATCGCGAGTTCAGTCTGGGG---TCGGACAAGCGCGACAACT 1522

Qy       1062 ------------------------------------------------------------ 1062

Db       1521 ATGTATGTCACTCAATCATTTTCCCAGGCTTTCCCTTCTAAATTATATTTATACATATAT 1462

Qy       1063 ----------GluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerL 1079
                        :::|||::::::   :::||||||:::||||||    ||||||    ||||
Db       1461 TTTTCGGTTATCAGAAAATGTGGGCGCTCTGCGATCTTGCCATGTACATTATCGACTCCA 1402

Qy       1079 ys--------------------SerThrThrTyrSerLeuGluSerProLysAspProV 1092
              ||                    ::::::|||:::|||:::
Db       1401 AGTTCAGTCCATTTGATGGCAACACGAGCACCTTTTCCATGCCGCTG-----------G 1354

Qy       1092 alLeuProAlaArgPhePheThrGlnProAsp---LysAsnPheSerAsnThrLysAsnT 1111
                ||||||    ::::::    :::|||            ||||||    |||            |
Db       1353 CTTTGCCAGAAATGTATTATAAAGAGCCTGCCGTTGCGAATTTCCAAAACAATGACGTCT 1294

Qy       1111 yrLeuProProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValL 1131
              ||:::|||    ::::::             :::|||    |||          |||:::
Db       1293 ATATACCGCTGGACGTG---------TATACGCTGGGAGCCAAATCCACGAGCAAAGCTG 1243

Qy       1131 euGlyAlaValAsnLysProLeu---SerSerAlaGlyLysGlnSerGlnThrLysSerS 1150
                          |||       :::   ||||||:::|||    |||          ::::::|
Db       1242 CCGCAACAGC-AATGACAACGTCGCGAGCAGCAGTGGCTC-CAAAGAGACCGGCGGAACA 1185

Qy       1150 erArg--------------------------MetGluThrValSerAsnAlaSerSerS 1161
              ||:::                          :::    ::::::    |||       :::
Db       1184 GTCAATCATGGATGATGAAAATCCGCAGGTGAGTGAGGTCCTTGGAAAATACTGAGAATG 1125

Qy       1161 erSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspH 1181
                ||||||                            :::       :::   |||:::
Db       1124 AATCTAACTAGTGTGTGAAATTTGATTTCAGGAGAACAATCTGTTCGACAACATAC---- 1069

Qy       1181 isSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
                        |||       |||:::   ||||||    |||    :::::::::    :::
Db       1068 ----------GAGCGGCTGACACTACGGAGCCCATGGCCAAACGAACGCGCGCAGGAG 1021
```

RESULT 11
ABK35358
ID ABK35358 standard; cDNA; 2496 BP.
XX
AC ABK35358;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human cDNA encoding secreted protein #496.
XX
KW Human; secreted protein; gene; ss; nutritional supplement; haemophilia;
KW viral infection; bacterial infection; fungal infection; diabetes; asthma;
KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;
KW autoimmune thyroiditis; allergic reaction; neurodegenerative disease;
KW Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;
KW coagulation disorder; inflammatory disorder; Crohn's disease; incision;
KW tissue regeneration; wound healing; burn; haematopoiesis;
KW myeloid cell deficiency; lymphoid cell deficiency.
XX
OS Homo sapiens.
XX
PN WO200177288-A2.
XX
PD 18-OCT-2001.
XX
PF 29-MAR-2001; 2001WO-US10224.
XX
PR 06-APR-2000; 2000US-195582P.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gulukota K, Graham JR;
XX
DR WPI; 2002-179321/23.
XX
PT Five hundred and ninety two polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for
PT treating immune deficiencies and disorders such as autoimmune disorders
PT -
XX
PS Claim 1; Page 328-329; 372pp; English.
XX
CC The invention relates to 592 polynucleotides which have been derived from
CC a variety of human tissue sources and which encode novel secreted
CC proteins. The polynucleotides can be used as probes for the
CC identification and isolation of full length cDNA and genomic DNA. The
CC polynucleotides and proteins can also be used as nutritional supplements.
CC The proteins are useful in the treatment of various immune deficiencies
CC and disorders such as viral infections, bacterial infections, fungal
CC infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple
CC sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions
CC and conditions (e.g. asthma). They are also useful for treating
CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
CC useful for tissue regeneration, for wound healing and in the treatment of
CC burns, incisions and ulcers. The proteins are also useful for regulating
CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies.
CC Sequences ABK34863-ABK35454 represent polynucleotides of the invention.
XX
SQ Sequence 2496 BP; 822 A; 430 C; 491 G; 753 T; 0 other;

Alignment Scores:

Pred. No.:	1.3e-88	Length:	2496
Score:	1409.50	Matches:	300
Percent Similarity:	65.96%	Conservative:	76
Best Local Similarity:	52.63%	Mismatches:	123
Query Match:	19.60%	Indels:	72
DB:	24	Gaps:	8

US-09-512-581B-2 (1-1391) x ABK35358 (1-2496)

```
Qy        717 LysSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIle 736
              |||:::|||:::|||    |||    |||||||||    |||:::||||||||||||||||||
Db          4 AAAGCAAAGAGGGGTACTCCACACCAAGCAAAACAGGCTGTGCACTGTATACACGCCATA 63

Qy        737 PheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAsp 756
              |||::::::||||||    |||    |||||||||||||||
Db         64 TTCACAAATAAAGAAGTCCAGCTTGCACAGATTTTTGAGTCA------------------ 105

Qy        757 ProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeu 776

Db        105 ------------------------------------------------------------ 105

Qy        777 AlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAsp 796

Db        105 ------------------------------------------------------------ 105
```

```
Qy    797 LeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGlu 816
Db    106 --------------------ACAGGTGAAAAGAATGGAAAACTGTGGTCTCCAGATGAA 144

Qy    817 GluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeu 836
Db    145 GAGGTTTCCCCTGAAGTACTAGCAAAGGTACAGGCAATTAAACTTCTGGTAAGGTGGCTG 204

Qy    837 LeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThr 856
Db    205 TTGGGTATGAAAAACAACCAGTCTAAATCTGCCAATTCAACCCTTCGGTTATTATCAGCG 264

Qy    857 IleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSer 876
Db    265 ATGTTGGTTAGTGAGGGTGACCTGACAGAGCAAAAGAGGATCAGTAAATCTGATATGTCT 324

Qy    877 ArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHis 896
Db    325 CGCTTGCGATTAGCTGCTGGTAGTGCCATAATGAAGCTTGCTCAGGAACCTTGTTACCAT 384

Qy    897 GluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyr 916
Db    385 GAAATTATTACCCCAGAACAGTTTCAGCTCTGTGCACTTGTTATTAATGATGAGTGTTAC 444

Qy    917 GlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuPro 936
Db    445 CAAGTAAGGCAGATATTTGCTCAGAAGCTGCATAAGGCACTTGTGAAGTTACTGCTCCCA 504

Qy    937 LeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAla 956
Db    505 TTGGAGTATATGGCGATCTTTGCCTTGTGTGCCAAAGATCCTGTGAAGGAGAGAAGAGCA 564

Qy    957 HisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHis 976
Db    565 CACGCACGACAATGTTTACTGAAAAATATCAGTATACGCAGGGAATACATTAAGCAGAAT 624

Qy    977 AlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIle 996
Db    625 CCTATGGCTACTGAGAAATTATTATCACTGTTGCCTGAATATGTAGTTCCATACATGATT 684

Qy    997 HisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAsp 1016
Db    685 CACCTGCTAGCCCATGATCCAGATTTTACAAGATCACAAGATGTTGATCAGCTTCGTGAT 744

Qy   1017 ValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSer 1036
Db    745 ATCAAAGAGTGCCTATGGTTCATGCTTGAAGTTTTAATGACAAAGAATGAAAACAATAGC 804

Qy   1037 HisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPro 1056
Db    805 CATGCCTTTATGAAGAAGATGGCAGAGAACATCAAGTTAACCAGAGATGCCCAGTCTCCA 864

Qy   1057 AspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIle 1076
Db    865 GATGAATCCAAGACAAATGAAAAACTGTATACAGTATGTGATGTGGCTCTCTGTGTTATA 924

Qy   1077 MetSerLysSerThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArg 1096
Db    925 AATAGTAAAAGTGCTTTGT--GCAATGCAGATTCCCAAAGGACCCAGCCTCCAATGAAAT 982

Qy   1097 PhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMet 1116
Db    983 TTTTACACA---CCTGAAAAGGACTTCTGTAACGATAAGAGTTATATTTCAGAAGAGACA 1039

Qy   1117 LysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLys 1136
Db   1040 AGAGTACTTCTGTTAACAGGAAAGCCAAAGCCTGCTGGAGTACTAGGTGCAGTAAATAAG 1099

Qy   1137 ProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSer 1156
Db   1100 CCTTTATCAGCAACGGGAAGGAAACCCTATGTTAGAAGCACTGGCACTGAGACTGGAAGC 1159

Qy   1157 ------AsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArg--- 1173
Db   1160 AATATTAATGTAAATTCAGAGCTGAACCCTTCAACCGGAAATCGATCAAGGGAACAGAGT 1219

Qy   1174 LeuAspSerSerGluMetAspHisSerGluAsnGluAsp------TyrThrMetSerSer 1191
Db   1220 TCAGAGGCAGCAGAAACTGGAGTTAGTGAAAATGAAGAGAACCCTGTGAGGATTATTTCA 1279

Qy   1192 ProLeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeu 1211
Db   1280 GTCACACCTGTAAAGAATATTGACCCAGTAAAGAATAAG------GTAAAAATGCATTTG 1333

Qy   1212 GluLysProArgGlyArgLysLysThrProValThrGluGlnGlu-GluLysLeuGlyMe 1231
Db   1334 CAAAGGGAGAAAATGAAGGCCAAACAGAAGCAGGCTCCAGCTTCTGCAAAAACTTGGATT 1393

Qy   1231 tAspAspLeu-----------ThrLysLeuValGlnGluGln--------------Ly 1242
Db   1394 CACAAATGTCCCTGAACAGAAAATGAAGCTCACTTCAGAACACACACTCTCTGCCTTGAA 1453

Qy   1242 sProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluGl 1262
Db   1454 AACTAAAGAGACTATTACTTCCTTTTCACATGACCACAAGTCCTCTGATGGAAATGTACA 1513

Qy   1262 nGlnTrpProGluGluLysArgLeuLys 1271
Db   1514 GCAGAAACTCTTGAGAGAGAGGCTAAAA 1541
```

RESULT 12
ABS56684

```
ID   ABS56684 standard; cDNA; 1104 BP.
XX
AC   ABS56684;
XX
DT   23-JAN-2003  (first entry)
XX
DE   Euchromosome fragile intelligence delay syndrome protein 20.02 cDNA.
XX
KW   Euchromosome fragile intelligence delay syndrome protein 20.02; human;
KW   euchromosome fragile intelligence delay syndrome; dementia; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             184..732
FT                   /*tag=  a
FT                   /product= "euchromosome fragile intelligence delay
FT                            syndrome-associated protein, 20.02"
XX
PN   CN1351041-A.
XX
PD   29-MAY-2002.
XX
PF   26-OCT-2000; 2000CN-0125797.
XX
PR   26-OCT-2000; 2000CN-0125797.
XX
PA   (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX
PI   Mao Y,  Xie Y;
XX
DR   WPI; 2002-637117/69.
DR   P-PSDB; ABB84560.
XX
PT   Human euchromosome fragile intelligence delay syndrome associated
PT   protein 20.02 polypeptide, used to treat e.g. dementia -
XX
PS   Claim 6; Page 29-30 (disclosure); 33pp; Chinese.
XX
CC   This invention describes the novel human euchromosome fragile
CC   intelligence delay syndrome associated protein, 20.02. The polypeptide is
CC   used in treating diseases such as euchromosome fragile intelligence delay
CC   syndrome, and dementia. This sequence encodes the human euchromosome
CC   fragile intelligence delay syndrome-associated protein, 20.02 described
```

```
CC   in the disclosure of the invention.
XX
SQ   Sequence 1104 BP; 362 A; 188 C; 228 G; 326 T; 0 other;

Alignment Scores:
Pred. No.:              3.13e-58       Length:        1104
Score:                  967.00         Matches:       183
Percent Similarity:     90.56%         Conservative:  28
Best Local Similarity:  78.54%         Mismatches:    22
Query Match:            13.44%         Indels:        1
DB:                     24             Gaps:          0

US-09-512-581B-2 (1-1391) x ABS56684 (1-1104)

Qy        348 AlaLysAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleArg 367
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GCGAAGGATCTCACAGAATATTTAAAGGTTAGATCACATGATCCAGAAGAAGCTATTCGT 61

Qy        368 HisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeuLeuValAsnAsp 387
              |||||||||||||||:::|||:::||||||||||||:::|||:::   ||||||||||||
Db         62 CATGATGTCATTGTTACTATAATAACAGCTGCCAAGAGGGACCTGCC-TTAGTAAATGAT 120

Qy        388 HisLeuLeuAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLysGlu 407
                 ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAGCTGCTTGGCTTTGTAAGGGAAAGAACACTGGATAAACGGTGGCGAGTAAGAAAAGAA 180

Qy        408 AlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyLys 427
              |||||||||||||||||||||:::||||||||||||   |||         |||||||||
Db        181 GCTATGATGGGTCTGGCTCAGCTTTATAAGAAATACTGTCTTCATGGTGAAGCAGGAAAG 240

Qy        428 AspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsn 447
              :::||||||::::::::::::|||||||||||||||||||||||||||||||||||||||
Db        241 GAAGCTGCAGAGAAAGTCAGCTGGATAAAGGACAAACTTCTGCATATTTATTATCAGAAC 300

Qy        448 SerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAsn 467
              ||||||||||||:::||||||||||||:::||||||||||||||||||:::|||||||||
Db        301 AGCATTGACGACAAATTGTTGGTAGAGAAAATCTTTGCTCAGTATCTTGTCCCCCACAAC 360

Qy        468 LeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsn 487
              |||||||||   |||||||||||||||||||||||||||||||||:::||||||   |||
Db        361 CTGGAAACAGAAGAGAGAATGAAATGCTTATATTACTTATATGCTAGTTTGGATCCAAAT 420

Qy        488 AlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLys 507
              |||||||||||||||||||||||||||||||||||||||:::||||||      |||:::
Db        421 GCTGTAAAAGCTCTCAACGAAATGTGGAAGTGTCAGAACATGCTTCGGAGCCATGTACGC 480

Qy        508 AspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSer 527
              :::||||||||||||   |||||||||   ::::::|||:::      |||:::|||
Db        481 GAACTATTGGATTTGCACAAGCAGCCTACATCAGAGGCTAACTGTTCTGCCATGTTTGGA 540

Qy        528 LysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLys 547
              |||:::|||   |||   :::|||||||||||||||||||||||||||||||||:::|||
Db        541 AAACTGATGACCATAGCAAAGAATTTGCCTGACCCCGGGAAAGCACAAGATTTTGTGAAG 600

Qy        548 LysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuVal 567
              ||||||   |||||||||   ||||||||||||:::|||   |||||||||:::|||:::
Db        601 AAATTTAACCAGGTTCTCGGCGATGATGAGAAACTTCGGTCTCAGTTGGAGTTATTAATT 660

Qy        568 SerProThrCysSerCysLysGlnAlaGluGlyCysVal 580
              |||||||||||||||||||||||||||:::   ||||||
Db        661 AGCCCAACCTGTTCTTGCAAACAAGCAGATATTTGTGTG 699

RESULT 13
AAK85948
ID   AAK85948 standard; DNA; 7943 BP.
XX
AC   AAK85948;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40760.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
```

```
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 40760; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 7943 BP; 2567 A; 1175 C; 1468 G; 2733 T; 0 other;

Alignment Scores:
Pred. No.:              2.6e-53       Length:        7943
Score:                  912.50        Matches:       180
Percent Similarity:     85.31%        Conservative:  0
Best Local Similarity:  85.31%        Mismatches:    0
Query Match:            12.69%        Indels:        31
DB:                     22            Gaps:          1

US-09-512-581B-2 (1-1391) x AAK85948 (1-7943)

Qy       1212 GluLysProArgGlyArgLysLysThrProValThrGluGlnGluGluLysLeuGlyMet 1231
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GAGAAGCCTAGAGGCAGGAAAAAAACGCCCGTCACAGAACAGGAGGAGAAATTAGGTATG 61

Qy       1232 AspAspLeuThrLysLeuValGlnGluGlnLysProLysGlySerGlnArgSerArgLys 1251
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GATGACTTGACTAAGTTGGTACAGGAACAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAA 121

Qy       1252 ArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrpProGluGluLysArgLeuLys 1271
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 AGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAA 181

Qy       1272 GluAspIleLeuGluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGly 1291
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 GAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGC 241

Qy       1292 ArgProProLysProLeuGlyGlyGlyThrProLysGluGluProThrMetLysThrSer 1311
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 CGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAGCCAACAATGAAAACTTCT 301

Qy       1312 LysLysGlySerLysLysLysSerGlyProProAlaProGluGluGluGluGluGluGlu 1331
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 AAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAA 361

Qy       1332 ArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysGlnHisArgValSerArgArg 1351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 AGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAACAGCACCGAGTGTCAAGGAGA 421

Qy       1352 AlaGlnGln--------------------------------------------------- 1354
              |||||||||
```

```
Db        422 GCACAGCAGAGGTAAGCATGTGTAACTCTAAACTGCATCTGTTTCGTTACTATATTATAA 481

Qy       1355 ----------------------------------------ArgAlaGluSerProGlu 1360
                                                      ||||||||||||||||||
Db        482 ATCATAATTTGATGCTATCCACATTTGGGTCTTCCCCAAAGCAGAGCAGAATCTCCTGAA 541

Qy       1361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 601

Qy       1381 ThrProSerProSerGlnProLysLysAsnVal 1391
              |||||||||||||||||||||||||||||||||
Db        602 ACGCCATCACCATCACAACCAAAAAAAAATGTG 634

RESULT 14
AAS44727
ID   AAS44727 standard; DNA; 714 BP.
XX
AC   AAS44727;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human full-length polynucleotide sequence #152.
XX
KW   Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW   mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW   cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW   nervous system disorder; inflammatory disorder; cell differentiation; ds;
KW   angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW   genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW   cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;
KW   antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW   neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;
KW   immunostimulant; analgesic; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200164834-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04926.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
PR   17-JUN-2000; 2000US-0597707.
PR   14-JUL-2000; 2000US-0616807.
PR   19-SEP-2000; 2000US-0664641.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Wehrman T,  Wang J,  Ma Y,  Wang D,  Chen R,  Xu C;
PI   Drmanac R;
XX
DR   WPI; 2001-589862/66.
DR   P-PSDB; AAU27827.
XX
PT   Novel polypeptides and nucleic acids obtained from cDNA libraries
PT   prepared from various human tissues, for diagnosis, treatment of
PT   cancer, neurological, inflammatory disorders and for use in arrays for
PT   detection  -
XX
PS   Claim 1; SEQ ID No 152; 153pp; English.
XX
CC   Sequences AAS44576-AAS44919 represent full-length polynucleotides and
CC   contig polynucleotides encoding polypeptides of the invention. The DNA
CC   and protein sequences are useful for the treatment, diagnosis and
CC   prevention of various types of disorder in a mammalian subject such as a
CC   human, dog, monkey, mouse, hamster or rat. The disorders include cancers
CC   such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such
CC   as multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC   diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC   disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC   chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC   Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC   disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC   bowel disease. The sequences exhibit activity relating to angiogenesis,
CC   cell proliferation, cell differentiation, stem cell growth factor,
CC   activin or inhibin. Therefore, they can be used to manipulate stem cells
CC   in culture to give rise to neuroepithelial cells that can be used to
CC   augment or replace cells damaged by illness, accidental damage or genetic
CC   disorders. The sequences may also be used for regeneration of bone,
CC   cartilage, tendons and ligaments and in tissue repair and burn healing.
CC   Note: Some sequences for this patent did not form part of the printed
CC   specification, but were obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 714 BP; 293 A; 131 C; 178 G; 112 T; 0 other;

Alignment Scores:
Pred. No.:              2.68e-51        Length:        714
Score:                  864.00          Matches:       166
Percent Similarity:     100.00%         Conservative:  0
Best Local Similarity:  100.00%         Mismatches:    0
Query Match:            12.01%          Indels:        0
DB:                     22              Gaps:          0

US-09-512-581B-2 (1-1391) x AAS44727 (1-714)

Qy       1226 GluGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGluGlnLysProLysGly 1245
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 GAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAACAGAAACCTAAAGGC 62

Qy       1246 SerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrpPro 1265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 AGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGTGGCCT 122

Qy       1266 GluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsnSerProPro 1285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 GAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCA 182

Qy       1286 LysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGlyThrProLysGluGlu 1305
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 AAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAG 242

Qy       1306 ProThrMetLysThrSerLysLysGlySerLysLysLysSerGlyProProAlaProGlu 1325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 CCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAG 302

Qy       1326 GluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLysSerLysSerLysGln 1345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 GAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAACAG 362

Qy       1346 HisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGluSerSerAlaIleGlu 1365
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 CACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAATCTAGTGCAATTGAA 422

Qy       1366 SerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLysThrProSerProSer 1385
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 TCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAAACGCCATCACCATCA 482

Qy       1386 GlnProLysLysAsnVal 1391
              ||||||||||||||||||
Db        483 CAACCAAAAAAAAATGTC 500

RESULT 15
AAF18338
ID   AAF18338 standard; DNA; 1217 BP.
XX
AC   AAF18338;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Lung cancer associated polynucleotide sequence SEQ ID 357.
XX
```

```
KW   Human; lung cancer associated protein; neuroprotective; cytostatic;
KW   cardioactive; immunomodulatory; muscular active; vulnerary;
KW   gastrointestinal; nephrotropic; antiinfective; gynecological;
KW   antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW   proliferative disorder; wound healing; infectious disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200055180-A2.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05918.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Ruben SM;
XX
DR   WPI; 2000-587514/55.
DR   P-PSDB; AAB58462.
XX
PT   Lung cancer associated gene sequences, referred to as lung cancer
PT   antigens, useful for treatment, prevention, and diagnosis of disorders
PT   such as lung cancer -
XX
PS   Claim 1; Page 815; 1425pp; English.
XX
CC   Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC   associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC   associated proteins and polynucleotide sequences, their agonists, and
CC   antagonists may have neuroprotective; cytostatic; cardioactive;
CC   immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC   general; nephrotropic; antiinfective; gynecological; or antibacterial
CC   activity. The invention also includes antibodies specific for the
CC   protein or polynucleotide sequences. The lung cancer associated
CC   polynucleotide sequences may be used for detection of lung cancer,
CC   chromosome identification, as chromosome markers, and for numerous other
CC   diagnostic or research purposes. The proteins may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders. The proteins may also be used in the treatment of wounds and
CC   infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC   peptide AAB58549 are used in the course of the invention for the
CC   identification and characterisation of the polynucleotide and protein
CC   sequences.
XX
SQ   Sequence 1217 BP; 445 A; 232 C; 270 G; 263 T; 7 other;

Alignment Scores:
Pred. No.:              6.38e-44      Length:        1217
Score:                  762.50        Matches:       185
Percent Similarity:     56.88%        Conservative:  63
Best Local Similarity:  42.43%        Mismatches:    121
Query Match:            10.60%        Indels:        68
DB:                     21            Gaps:          15

US-09-512-581B-2 (1-1391) x AAF18338 (1-1217)

Qy        965 AsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeu 984
              ||||||::::::|||||||||||||:::||||||:::          :::|||||||||||
Db          2 AATATCAGTATACGCAGGGAATACATTAAGCAGAATCCTATGGCTACTGAGAAATTATTA 61

Qy        985 SerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAsp 1004
              ||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db         62 TCACTGTTGCCTGAATATGTAGTTCCATACATGATTCACCTGCTAGCCCATGATCCAGAT 121

Qy       1005 TyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheVal 1024
              :::   :::   ||||||::::::||||||:::|||:::||||||||||||||||||:::
Db        122 TTTACAAGATCACAAGATGTTGATCAGCTTCGTGATATCAAAGAGTGCCTATGGTTCATG 181

Qy       1025 LeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetVal 1044
              ||||||:::||||||   |||||||||||||||||||||||||||::::::||||||
Db        182 CTTGAAGTTTTAATGACAAAGAATGAAAACAATAGCCATGCCTTTATGAAGAAGATGGCA 241

Qy       1045 GluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLys 1064
              ||||||||||||   |||:::|||||||||   ||||||::::::|||   |||||||||
Db        242 GAGAACATCAAGTTAACCAGAGATGCCCAGTCTCCAGATGAATCCAAGACAAATGAAAAA 301

Qy       1065 LeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSer 1084
              ||||||||||||||||||||||||:::   :::|||   |||||||||         :::
Db        302 CTGTATACAGTATGTGATGTGGCTCTCTGTGTTATAAATAGTAAAAGTGCTTTGTGCAAT 361

Qy       1085 LeuGluSerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsn 1104
                 :::||||||||||||||||||||||||   :::|||||||||||||||:::|||:::
Db        362 GCAGATTCACCAAAGGACCCAGTCCTCCCAATGAAATTTTTACACAACCTGAAAAGGAC 421

Qy       1105 PheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePheThrProGlyLys 1124
              |||   |||   |||:::|||:::         |||   :::            ||||||
Db        422 TTCTGTAACGATAAGAGTTATATTTCAGAAGAGACAAGAGTACTTCTGTTAACAGGAAAG 481

Qy       1125 ProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGln 1144
              ||||||   |||||||||||||||||||||||||||||||||:::   |||::::::
Db        482 CCAAAGCCTGCTGGAGTACTAGGTGCAGTAAATAAGCCTTTATCAGCAACGGGAAGGAAA 541

Qy       1145 SerGlnThrLysSerSerArgMetGluThrValSer------AsnAlaSerSerSerSer 1162
                 :::|||:::      ||||||   |||         |||   :::|||
Db        542 CCCTATGTTAGAAGCACTGGCACTGAGACTGGAAGCAATATTAATGTAAATTCAGAGCTG 601

Qy       1163 AsnProSerSerProGlyArgIleLysGlyArg---LeuAspSerSerGluMetAspHis 1181
              ||||||||||:::      |||   :::   :::         :::::::::|||
Db        602 AACCCTTCAACCGGAAATCGATCAAGGGAACAGAGTTCAGAGGCAGCAGAAACTGGAGTT 661

Qy       1182 SerGluAsnGluAsp------TyrThrMetSerSerProLeuProGlyLysLysSerAsp 1199
              ||||||||||||:::            :::   |||      |||   |||      |||
Db        662 AGTGAAAATGAAGAGAACCCTGTGAGGATTATTTCAGTCACACCTGTAAAGAATATTGAC 721

Qy       1200 LysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLys 1219

Db        721 ------------------------------------------------------------ 721

Qy       1220 ThrProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuValGln 1239
                 ||||||      :::|||      :::      |||      |||
Db        722 ---CCAGTAAAGAATAAGGAA------ATTAATTCTGATCAGGCTACC------------ 760

Qy       1240 GluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSer 1259
                 |||         |||   |||   :::||||||      ||||||:::      :::
Db        761 ---CAGGGCAACATCAGCAGTGACCGAGGAAAGAAAAGAACAGTAACAGCAGCTGGTGCA 817

Qy       1260 AspGlu---GlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGlu 1278
              :::      ||||||      :::||||||
Db        818 GAGAATATCCAACAAAAAACAGATGAGAAA--------------------------GTA 850

Qy       1279 AspGluGlnAsnSerPro---ProLysLysGlyLysArgGlyArgProProLysProLeu 1297
              ||||||         |||   |||   |||   :::|||||||||   ||||||
Db        851 GATGAATCGGGACCTCCCGCCCCTTCCAAACCCAGGAGAGGACGTCGACCCAAGTCTGAA 910

Qy       1298 Gly---GlyGlyThrProLysGluGluProThrMetLysThrSerLysLysGlySerLys 1316
                    |||         |||   :::            |||         ||||||   |||
Db        911 TCTCAGGGCAATGCTACCAAAAATGATGATCTAAATAAACCTATTAACAAGGGAAGGAAG 970

Qy       1317 LysLysSer---GlyProProAlaProGluGluGluGlu--------------------- 1328
              :::   :::   |||         :::|||         |||
Db        971 AGAGCTGCAGTGGGTCAGGAGAGCCCTGGGGGTTTGGAAGCAGGTAATGCCAAAGCACCC 1030

Qy       1329 ------GluGluGluArgGlnSerGlyAsnThrGluGlnLys------------SerLys 1342
                    ::::::      |||   |||:::|||   :::|||            :::|||
Db       1031 AAACTGCAAGATTT-AGCCAAAAAGGCAGCACCAGCAGAAAGACMAATTGACTTACAAAG 1089

Qy       1343 SerLysGlnHis------ArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
                 |||   |||         |||         :::|||:::|||:::
Db       1090 GTRAAAATGCATTTGCAAAGGGAGAAAATGAAGGCCAAACAGAAGCAGGCTNCCAGYTTN 1149

Qy       1361 SerSerAlaIle------------------GluSerThrGlnSerThr 1370
```

```
                      :::                       |||   |||   ||||||
Db        1150 TGCAAAANCTTGGATTACAATGKCCTGACAGAAATGACTTATTCAACC 1197
```

Search completed: September 24, 2003, 20:46:25
Job time : 949 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

```
Run on:          September 24, 2003, 19:50:39 ; Search time 175 Seconds
                                                (without alignments)
                                                2051.150 Million cell updates/sec

Title:           US-09-512-581B-2
Perfect score:   7193
Sequence:        1 MAHSKTRTNDGKITYPPGVK..........QKGRGRPSKTPSPSQPKKNV 1391

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters:      830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_23:*
                 1:  sp_archea:*
                 2:  sp_bacteria:*
                 3:  sp_fungi:*
                 4:  sp_human:*
                 5:  sp_invertebrate:*
                 6:  sp_mammal:*
                 7:  sp_mhc:*
                 8:  sp_organelle:*
                 9:  sp_phage:*
                 10:  sp_plant:*
                 11:  sp_rodent:*
                 12:  sp_virus:*
                 13:  sp_vertebrate:*
                 14:  sp_unclassified:*
                 15:  sp_rvirus:*
                 16:  sp_bacteriap:*
                 17:  sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7155	99.5	1447	4	Q9NTI5	Q9nti5 homo sapien
2	7155	99.5	1483	4	Q9Y2I5	Q9y2i5 homo sapien
3	7101	98.7	1391	4	Q9Y451	Q9y451 homo sapien
4	4668.5	64.9	1297	4	Q8NG14	Q8ng14 homo sapien
5	3787	52.6	824	11	Q8BLH6	Q8blh6 mus musculu
6	3456	48.0	701	11	Q8BJ18	Q8bj18 mus musculu
7	2698.5	37.5	851	4	Q9Y4D4	Q9y4d4 homo sapien
8	2517	35.0	529	4	Q8IXT6	Q8ixt6 homo sapien
9	2487	34.6	479	4	Q96KV4	Q96kv4 homo sapien
10	2476.5	34.4	536	4	Q96KV3	Q96kv3 homo sapien
11	2474	34.4	600	4	Q8N7J4	Q8n7j4 homo sapien
12	2317	32.2	584	11	Q8VDS0	Q8vds0 mus musculu
13	1982.5	27.6	1218	5	Q9V646	Q9v646 drosophila
14	833.5	11.6	1205	3	Q9HFF5	Q9hff5 schizosacch
15	808	11.2	1638	10	Q9FIL0	Q9fil0 arabidopsis
16	807	11.2	1579	5	Q9XTF6	Q9xtf6 caenorhabdi
17	791	11.0	198	11	Q9D337	Q9d337 mus musculu
18	747.5	10.4	1596	3	Q9UVY6	Q9uvy6 sordaria ma
19	656	9.1	125	4	Q9H5N8	Q9h5n8 homo sapien
20	640.5	8.9	333	4	Q96DB6	Q96db6 homo sapien
21	632.5	8.8	1506	3	O94076	O94076 emericella
22	602	8.4	192	11	Q8BX77	Q8bx77 mus musculu
23	572.5	8.0	1378	10	Q8S1N0	Q8s1n0 oryza sativ
24	571.5	7.9	1303	10	Q9CAP7	Q9cap7 arabidopsis
25	487	6.8	415	5	Q95U25	Q95u25 drosophila
26	423.5	5.9	1033	5	Q8MMY4	Q8mmy4 dictyosteli
27	309.5	4.3	390	3	O94237	O94237 schizosacch
28	303	4.2	780	10	Q9SAI0	Q9sai0 arabidopsis
29	264	3.7	990	10	Q9S9P0	Q9s9p0 arabidopsis
30	256	3.6	873	10	Q8GUP3	Q8gup3 arabidopsis
31	248.5	3.5	2158	4	Q9Y6Y3	Q9y6y3 homo sapien
32	242.5	3.4	2265	4	Q9Y6Y4	Q9y6y4 homo sapien
33	235	3.3	1135	6	Q8HZW3	Q8hzw3 canis famil
34	234.5	3.3	852	10	Q9SZ55	Q9sz55 arabidopsis
35	234.5	3.3	3130	5	Q8IDX6	Q8idx6 plasmodium
36	232.5	3.2	3130	5	Q9BK46	Q9bk46 plasmodium
37	231.5	3.2	2253	13	P70012	P70012 xenopus lae
38	230	3.2	2501	5	Q9NCW7	Q9ncw7 drosophila
39	229	3.2	1723	2	Q9JMX8	Q9jmx8 helicobacte
40	228.5	3.2	18074	5	Q9I7U4	Q9i7u4 drosophila
41	225	3.1	1437	5	Q18033	Q18033 caenorhabdi
42	223.5	3.1	4717	3	O94248	O94248 schizosacch
43	223	3.1	2139	5	Q07569	Q07569 entamoeba h
44	223	3.1	3254	5	Q9BK45	Q9bk45 plasmodium
45	222.5	3.1	1819	16	Q9ZLV0	Q9zlv0 helicobacte

ALIGNMENTS

```
RESULT 1
Q9NTI5
ID   Q9NTI5     PRELIMINARY;      PRT;  1447 AA.
AC   Q9NTI5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rhodes S., Huckle E.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL137201; CAB69911.1; -.
DR   InterPro; IPR000637; AT_hook.
DR   SMART; SM00384; AT_hook; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   1447 AA;  164666 MW;  145C30308EA3EFD5 CRC64;

  Query Match            99.5%;  Score 7155;  DB 4;  Length 1447;
  Best Local Similarity  99.7%;  Pred. No. 0;
  Matches 1387;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLA 60

Qy         61 LHLASDFFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDT 120
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHLASDFFLKHPDKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDT 120

Qy        121 KSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDL 180

Qy        181 MSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTFFNQV 240
```

NPA

```
               |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db         181 MSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITNFFNQV 240

Qy         241 LMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMF 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 LMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMF 300

Qy         301 GAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSH 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSH 360

Qy         361 DPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYA 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 DPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYA 420

Qy         421 LQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYL 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 LQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYL 480

Qy         481 YATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITRNLPDPG 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 YATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITRNLPDPG 540

Qy         541 KAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEM 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 KAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEM 600

Qy         601 IKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTH 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 IKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTH 660

Qy         661 PISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKK 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 PISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKK 720

Qy         721 GPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQ 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 GPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQ 780

Qy         781 FAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMK 840
               |||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 FAAPLKSLVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMK 840

Qy         841 NNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIIT 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 NNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIIT 900

Qy         901 LEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQ 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 LEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQ 960

Qy         961 CLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKEC 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 CLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKEC 1020

Qy        1021 LWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKS 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1021 LWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKS 1080

Qy        1081 TTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSS 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 TTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSS 1140

Qy        1141 AGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDK 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1141 AGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDK 1200

Qy        1201 RDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESD.1260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 RDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESD 1260

Qy        1261 EQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSG 1320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1261 EQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSG 1320

Qy        1321 PPAPEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESPESSAIESTQSTPQKGRGRPSK 1380
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1321 PPAPEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESPESSAIESTQSTPQKGRGRPSK 1380

Qy        1381 TPSPSQPKKNV 1391
               |||||||||||
Db        1381 TPSPSQPKKNV 1391
```

```
RESULT 2
Q9Y2I5
ID   Q9Y2I5      PRELIMINARY;      PRT;  1483 AA.
AC   Q9Y2I5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   KIAA0979 protein (Fragment).
GN   KIAA0979.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99246063; PubMed=10231032;
RA   Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirosawa M.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIII.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:63-70(1999).
DR   EMBL; AB023196; BAA76823.2; -.
DR   InterPro; IPR000637; AT_hook.
DR   PRINTS; PR00929; ATHOOK.
FT   NON_TER       1       1
SQ   SEQUENCE   1483 AA;  168415 MW;  0E328407FD0EDD74 CRC64;

  Query Match             99.5%;  Score 7155;  DB 4;  Length 1483;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 1387;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy           1 MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLA 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          37 MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLA 96

Qy          61 LHLASDFFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDT 120
               |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db          97 LHLASDFFLKHPDKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDT 156

Qy         121 KSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDL 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         157 KSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDL 216

Qy         181 MSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTFFNQV 240
               |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db         217 MSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITNFFNQV 276

Qy         241 LMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMF 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         277 LMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMF 336

Qy         301 GAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSH 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         337 GAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSH 396

Qy         361 DPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYA 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         397 DPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYA 456

Qy         421 LQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYL 480
```

NPD

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          457 LQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYL 516

Qy          481 YATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITRNLPDPG 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          517 YATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITRNLPDPG 576

Qy          541 KAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEM 600
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          577 KAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEM 636

Qy          601 IKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTH 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          637 IKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTH 696

Qy          661 PISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKK 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          697 PISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKK 756

Qy          721 GPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQ 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          757 GPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQ 816

Qy          781 FAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMK 840
                |||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          817 FAAPLKSLVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMK 876

Qy          841 NNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIIT 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          877 NNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIIT 936

Qy          901 LEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQ 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          937 LEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQ 996

Qy          961 CLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKEC 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          997 CLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKEC 1056

Qy         1021 LWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKS 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1057 LWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKS 1116

Qy         1081 TTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSS 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1117 TTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSS 1176

Qy         1141 AGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDK 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1177 AGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDK 1236

Qy         1201 RDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESD 1260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1237 RDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESD 1296

Qy         1261 EQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSG 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1297 EQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSG 1356

Qy         1321 PPAPEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESPESSAIESTQSTPQKGRGRPSK 1380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1357 PPAPEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESPESSAIESTQSTPQKGRGRPSK 1416

Qy         1381 TPSPSQPKKNV 1391
                |||||||||||
Db         1417 TPSPSQPKKNV 1427
```

```
RESULT 3
Q9Y451
ID   Q9Y451      PRELIMINARY;      PRT;  1391 AA.
AC   Q9Y451;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)      NPA
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Androgen-induced prostate proliferative shutoff associated
DE   protein.
GN   AS3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Prostate;
RA   Geck P., Szelei J., Jimenez J., Soto A.M., Sonnenschein C.;
RT   "Androgen induced proliferative shutoff in prostate cancer cells.";
RL   Proc. Am. Assoc. Cancer Res 37:223-223(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Prostate;
RX   MEDLINE=99229875; PubMed=10215036;
RA   Geck P., Szelei J., Jimenez J., Sonnenschein C., Soto A.M.;
RT   "Early gene expression during androgen-induced inhibition of
RT   proliferation of prostate cancer cells: a new suppressor candidate on
RT   chromosome 13, in the BRCA2-Rb1 locus.";
RL   J. Steroid Biochem. Mol. Biol. 68:41-45(1999).
DR   EMBL; U95825; AAD22134.2; -.
DR   InterPro; IPR000637; AT_hook.
DR   SMART; SM00384; AT_hook; 1.
SQ   SEQUENCE   1391 AA;  158035 MW;  F58AEEE5AD6D9479 CRC64;

  Query Match             98.7%;  Score 7101;  DB 4;  Length 1391;
  Best Local Similarity   99.0%;  Pred. No. 0;
  Matches 1377;  Conservative    4;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          1 MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLA 60

Qy         61 LHLASDFFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDT 120
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHLASDFFLKHPDKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDT 120

Qy        121 KSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDL 180

Qy        181 MSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTFFNQV 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        181 MSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITNFFNQV 240

Qy        241 LMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMF 300
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        241 LMLGKTSISDLSEHVSDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMF 300

Qy        301 GAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSH 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSH 360

Qy        361 DPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYA 420
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db        361 DPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVGERTLDKRWRVRKEAMMGLAQIYKKYA 420

Qy        421 LQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYL 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYL 480

Qy        481 YATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITRNLPDPG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db        481 YATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITRSLPDPG 540

Qy        541 KAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 KAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEM 600
```

```
Qy        601 IKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTH 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTH 660

Qy        661 PISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKK 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKK 720

Qy        721 GPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQ 780
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        721 GPPRQAKYAIHCIHAIFSSKEAQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQ 780

Qy        781 FAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMK 840
              |||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 FAAPLKSLVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMK 840

Qy        841 NNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIIT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIIT 900

Qy        901 LEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQ 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQ 960

Qy        961 CLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKEC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKEC 1020

Qy       1021 LWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKS 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 LWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKS 1080

Qy       1081 TTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSS 1140
              |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db       1081 TTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPGMKSFFTPGKPKTTNVLGAVNKPLSS 1140

Qy       1141 AGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDK 1200
              ||||||||||||||| |||||||||||||||||||||||||||||||||||||||:|||
Db       1141 AGKQSQTKSSRMETVGNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKRSDK 1200

Qy       1201 RDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESD 1260
              ||||||||||||||||||||||||:||||||||||||||||:||||||||||||||||||
Db       1201 RDDSDLVRSELEKPRGRKKTPVTEREEKLGMDDLTKLVQEQRPKGSQRSRKRGHTASESD 1260

Qy       1261 EQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 EQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSG 1320

Qy       1321 PPAPEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESPESSAIESTQSTPQKGRGRPSK 1380
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db       1321 PPAPEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESSESSAIESTQSTPQKGRGRPSK 1380

Qy       1381 TPSPSQPKKNV 1391
              |||||||||||
Db       1381 TPSPSQPKKNV 1391
```

RESULT 4
Q8NG14
ID   Q8NG14   PRELIMINARY;   PRT; 1297 AA.
AC   Q8NG14;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   SCC-112.
GN   SCC-112.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kumar D., Patel S., Whiteside T.L., Kasid U.;
RT   "Identification and characterization of SCC-112, a novel cell cycle
RT   regulated gene in human cancer cells.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF294791; AAM82347.1; -.
SQ   SEQUENCE   1297 AA; 146610 MW; A42D13E68CA5808E CRC64;

Query Match 64.9%; Score 4668.5; DB 4; Length 1297;
Best Local Similarity 69.0%; Pred. No. 2.1e-261;
Matches 908; Conservative 160; Mismatches 199; Indels 49; Gaps 11;

```
Qy         31 MVRRLKMVVKTFMDMDQDSEEEKELYLNLALHLASDFFLKHPGKDVRLLVACCLADIFRI 90
              |::|||||||||||||||||:||: || |||||||:|||::| |||||||||||||||||
Db          1 MIKRLKMVVKTFMDMDQDSEDEKQQYLPLALHLASEFFLRNPNKDVRLLVACCLADIFRI 60

Qy         91 YAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSYNICFELEDS 150
              ||||||||| |||||||:|||||||||||||||||||||||||||:|||||||||||||
Db         61 YAPEAPYTSHDKLKDIFLFITRQLKGLEDTKSPQFNRYFYLLENLAWVKSYNICFELEDC 120

Qy        151 NEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSSIICEGDTVSQELLDTVLVNLVPAHKNL 210
              |||| ||:||||||||| ||:|| |||:||||||| ||| |:|||| ::|:||:||||||
Db        121 NEIFIQLFRTLFSVINNSHNKKVQMHMLDLMSSIIMEGDGVTQELLGSILINLIPAHKNL 180

Qy        211 NKQAYDLAKALLKRTAQAIEPYITTFFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLL 270
              |||::|||| ||||| | ||  |  ||||||:||::|:||||||||||| ||: || |||
Db        181 NKQSFDLAKVLLKRTVQTIEACIANFFNQVLVLGRSSVSDLSEHVFDLIQELFAIDPHLL 240

Qy        271 LSVLPQLEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIR 330
              |||:||||||||||| |||| ||:||||:||:|||:||:||:|||||:||||||||||:|
Db        241 LSVMPQLEFKLKSNDGEERLAVVRLLAKLFGSKDSDLATQNRPLWQCFLGRFNDIHVPVR 300

Qy        331 LECVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLL 390
              || |||||||||||||||||||||||||||||||||||||||:|:||||:|: |||| ||
Db        301 LESVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVTIITAAKRDLALVNDQLL 360

Qy        391 NFVRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSID 450
               |||||||||||||||||||||||:|||| |   |||:||::::||||||||||||||||
Db        361 GFVRERTLDKRWRVRKEAMMGLAQLYKKYCLHGEAGKEAAEKVSWIKDKLLHIYYQNSID 420

Qy        451 DRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLL 510
              |:||||:|||||:||||||| |||||||||||:|| ||||||||||||||:||  |::||
Db        421 DKLLVEKIFAQYLVPHNLETEERMKCLYYLYASLDPNAVKALNEMWKCQNMLRSHVRELL 480

Qy        511 DLIKQPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPT 570
              || ||| ::|:  |:| |:| | :|||||||||||:||| ||| ||||:| |||:|:|||
Db        481 DLHKQPTSEANCSAMFGKLMTIAKNLPDPGKAQDFVKKFNQVLGDDEKLRSQLELLISPT 540

Qy        571 CSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQVNKSID 630
              ||||||: ||||| :|| ||||||||||||:|||||||||||||:|:||||:| :||||:
Db        541 CSCKQADICVREIARKLANPKQPTNPFLEMVKFLLERIAPVHIDSEAISALVKLMNKSIE 600

Qy        631 GTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAALQ 690
              ||||||:|||  | |||:||||||||||||| |||||||:|||| ||:|:||||||||:|
Db        601 GTADDEEEGVSPDTAIRSGLELLKVLSFTHPTSFHSAETYESLLQCLRMEDDKVAEAAIQ 660

Qy        691 IFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEP 750
              ||:||| ||| | | ||| |:|:|| |:|:| | ||| |:|||||||::|| | ||||||
Db        661 IFRNTGHKIETDLPQIRSTLIPILHQKAKRGTPHQAKQAVHCIHAIFTNKEVQLAQIFEP 720

Qy        751 LHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPWKSWVATFIVKDLLMNDRLPGKKTTK 810
              | :||:     | ||||||::|||::|||||||:| || || |||||||||||  |:|  |
Db        721 LSRSLNADVPEQLITPLVSLGHISMLAPDQFASPMKSVVANFIVKDLLMNDRSTGEKNGK 780

Qy        811 LWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKI 870
              || |||||||| : |:||||::|||||||||| |||  ||||||: :| |:|||||| :|
Db        781 LWSPDEEVSPEVLAKVQAIKLLVRWLLGMKNNQSKSANSTLRLLSAMLVSEGDLTEQKRI 840

Qy        871 SKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGL 930
              || |||||||||||||:||||||||||||| ||:||||| ||||||||||:||||||| |
Db        841 SKSDMSRLRLAAGSAIMKLAQEPCYHEIITPEQFQLCALVINDECYQVRQIFAQKLHKAL 900

Qy        931 SRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEY 990
               :| |||||||| |||||||||||||||||||:|||::||||:||:   :||||||||||
Db        901 VKLLLPLEYMAIFALCAKDPVKERRAHARQCLLKNISIRREYIKQNPMATEKLLSLLPEY 960
```

```
Qy        991 VVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQT 1050
              |||| ||||||||: : ||::||:|:||||||:||:|| |||||||||::|| |||| |
Db        961 VVPYMIHLLAHDPDFTRSQDVDQLRDIKECLWFMLEVLMTKNENNSHAFMKKMAENIKLT 1020

Qy       1051 KDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKN 1110
              :||| ||::| ||||||||||: :| |||    : :||||||| :|||||:|:| | |:
Db       1021 RDAQSPDESKTNEKLYTVCDVALCVINSKSALCNADSPKDPVLPMKFFTQPEKDFCNDKS 1080

Qy       1111 YLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVS--NASSSSNPSSPG 1168
              |: | :        ||||  |||||||||: |::    :|:  || |  | :|  |||:
Db       1081 YISEETRVLLLTGKPKPAGVLGAVNKPLSATGRKPYVRSTGTETGSNINVNSELNPSTGN 1140

Qy       1169 RIKGR-LDSSEMDHSENED--YTMSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQ 1225
              | : :  :::|  ||||:     : |  | |  |                    ||  :
Db       1141 RSREQSSEAAETGVSENEENPVRIISVTPVKNID--------------------PVKNK 1179

Qy       1226 EEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESDE-QQWPEEKRLKEDILENEDEQNSP 1284
              |  :  |  |      |     | | :||  ||: ::  ||  :||          ||  |
Db       1180 E--INSDQAT-----QGNISSDRGKKRTVTAAGAENIQQKTDEK---------VDESGPP 1223

Qy       1285 -PKKGKRGRPPKPLG-GGTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEEERQSGNTE 1338
               | | :||| ||    |    | :    |    || |:      |  :|      ::|| :
Db       1224 APSKPRRGRRPKSESQGNATKNDDLNKPINKGRKR----AAVGQESPGGLEAGNAK 1275
```

```
RESULT 5
Q8BLH6
ID   Q8BLH6      PRELIMINARY;      PRT;   824 AA.
AC   Q8BLH6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   49J10.1.2 (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK045159; BAC32242.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   824 AA;  92412 MW;  4FED723F792C71EC CRC64;

  Query Match             52.6%;  Score 3787;  DB 11;  Length 824;
  Best Local Similarity   95.3%;  Pred. No. 1.1e-210;
  Matches  731;  Conservative   13;  Mismatches   23;  Indels    0;  Gaps    0;

Qy        625 VNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDEKV 684
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 VNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDEKV 60

Qy        685 AEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQF 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQF 120

Qy        745 AQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPWKSWVATFIVKDLLMNDRLP 804
              ||||||||||||||||||||||||||||||||||||||||| || ||||||||||||||||
Db        121 AQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPLKSLVATFIVKDLLMNDRLP 180

Qy        805 GKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDL 864
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDL 240

Qy        865 TEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQ 924
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQ 300

Qy        925 KLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLL 984
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        301 KLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNITVRREYLKQHAAVSEKLL 360

Qy        985 SLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMV 1044
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMV 420

Qy       1045 ENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKN 1104
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ENIKQTKDAQGPDDTKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKN 480

Qy       1105 FSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNP 1164
              |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        481 FSNTKNYLPPEMKSFFTPGKPKTANVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNP 540

Qy       1165 SSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTE 1224
              ||||||||||||||||||||||||||||||||||||:| ||||||||||| ||| |||:
Db        541 SSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKREDPDLVRSELEKPRSRKKAPVTD 600

Qy       1225 QEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESDEQQWPEEKRLKEDILENEDEQNSP 1284
               ||||||||||||||||||||| |||| |||:|||||||||||| ||::||||||||||
Db        601 PEEKLGMDDLTKLVQEQKPKGSQRGRKRGRTASDSDEQQWPEEKRHKEELLENEDEQNSP 660

Qy       1285 PKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEEERQSGNTEQKSKSK 1344
              |||||||||||||||| ||||||||||||:|||  ||  :::|||||| |||| |||||
Db        661 PKKGKRGRPPKPLGGGTSKEEPTMKTSKKGNKKKLVPPVVDDDEEEERQIGNTEHKSKSK 720

Qy       1345 QHRVSRRAQQRAESPESSAIESTQSTPQKGRGRPSKTPSPSQPKKNV 1391
              ||| |:|||||||||:||:||||||||||||||||||| |||||| | :
Db        721 QHRTSKRAQQRAESPETSAVESTQSTPQKGRGRPSKAPSPSQPPKKI 767
```

```
RESULT 6
Q8BJ18
ID   Q8BJ18      PRELIMINARY;      PRT;   701 AA.
AC   Q8BJ18;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   49J10.1.2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Thymus;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK041682; BAC31031.1; -.
SQ   SEQUENCE   701 AA;  78714 MW;  FAF6F0351AF7869B CRC64;

  Query Match             48.0%;  Score 3456;  DB 11;  Length 701;
  Best Local Similarity   98.8%;  Pred. No. 1.2e-191;
  Matches  671;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        530 MVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGN 589
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db          1 MVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEALVSPTCSCKQAEGCVREITKKLGN 60

Qy        590 PKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAG 120

Qy        650 LELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 LELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSA 180

Qy        710 LLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVT 769
Db        181 LLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVT 240

Qy        770 IGHIALLAPDQFAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAI 829
Db        241 IGHIALLAPDQFAAPLKSLVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAI 300

Qy        830 KMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKL 889
Db        301 KMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKL 360

Qy        890 AQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKD 949
Db        361 AQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKD 420

Qy        950 PVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQ 1009
Db        421 PVKERRAHARQCLVKNITVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQ 480

Qy       1010 DIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVC 1069
Db        481 DIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDTKMNEKLYTVC 540

Qy       1070 DVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTN 1129
Db        541 DVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTAN 600

Qy       1130 VLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTM 1189
Db        601 VLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTM 660

Qy       1190 SSPLPGKKSDKRDDSDLVR 1208
Db        661 SSPLPGKKSDKREDPDLVR 679
```

RESULT 7
Q9Y4D4
ID   Q9Y4D4   PRELIMINARY;   PRT;   851 AA.
AC   Q9Y4D4;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein KIAA0648 (Fragment).
GN   KIAA0648.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98403880; PubMed=9734811;
RA   Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
RA   Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
DR   EMBL; AB014548; BAA31623.1; -.
KW   Hypothetical protein.
FT   NON_TER   1   1
SQ   SEQUENCE   851 AA; 94791 MW; 4F00E954B8354807 CRC64;

Query Match 37.5%; Score 2698.5; DB 4; Length 851;
Best Local Similarity 62.1%; Pred. No. 9.8e-148;
Matches 540; Conservative 114; Mismatches 167; Indels 49; Gaps 11;

```
Qy        477 LYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITRNL 536
Db          1 LYYLYASLDPNAVKALNEMWKCQNMLRSHVRELLDLHKQPTSEANCSAMFGKLMTIAKNL 60

Qy        537 PDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNP 596
Db         61 PDPGKAQDFVKKFNQVLGDDEKLRSQLELLISPTCSCKQADICVREIARKLANPKQPTNP 120

Qy        597 FLEMIKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVL 656
Db        121 FLEMVKFLLERIAPVHIDSEAISALVKLMNKSIEGTADDEEEGVSPDTAIRSGLELLKVL 180

Qy        657 SFTHPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHH 716
Db        181 SFTHPTSFHSAETYESLLQCLRMEDDKVAEAAIQIFRNTGHKIETDLPQIRSTLIPILHQ 240

Qy        717 KSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALL 776
Db        241 KAKRGTPHQAKQAVHCIHAIFTNKEVQLAQIFEPLSRSLNADVPEQLITPLVSLGHISML 300

Qy        777 APDQFAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWL 836
Db        301 APDQFASPMKSVVANFIVKDLLMNDRSTGEKNGKLWSPDEEVSPEVLAKVQAIKLLVRWL 360

Qy        837 LGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYH 896
Db        361 LGMKNNQSKSANSTLRLLSAMLVSEGDLTEQKRISKSDMSRLRLAAGSAIMKLAQEPCYH 420

Qy        897 EIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRA 956
Db        421 EIITPEQFQLCALVINDECYQVRQIFAQKLHKALVKLLLPLEYMAIFALCAKDPVKERRA 480

Qy        957 HARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKD 1016
Db        481 HARQCLLKNISIRREYIKQNPMATEKLLSLLPEYVVPYMIHLLAHDPDFTRSQDVDQLRD 540

Qy       1017 VKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNII 1076
Db        541 IKECLWFMLEVLMTKNENNSHAFMKKMAENIKLTRDAQSPDESKTNEKLYTVCDVALCVI 600

Qy       1077 MSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNK 1136
Db        601 NSKSALCNADSPKDPVLPMKFFTQPEKDFCNDKSYISEETRVLLLTGKPKPAGVLGAVNK 660

Qy       1137 PLSSAGKQSQTKSSRMETVS--NASSSSNPSSPGRIKGR-LDSSEMDHSENED--YTMSS 1191
Db        661 PLSATGRKPYVRSTGTETGSNINVNSELNPSTGNRSREQSSEAAETGVSENEENPVRIIS 720

Qy       1192 PLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRK 1251
Db        721 VTPVKNID--------------------PVKNKE--INSDQAT-----QGNISSDRGKK 752

Qy       1252 RGHTASESDE-QQWPEEKRLKEDILENEDEQNSP-PKKGKRGRPPKPLG-GGTPKEEPTM 1308
Db        753 RTVTAAGAENIQQKTDEK---------VDESGPPAPSKPRRGRRPKSESQGNATKNDDLN 803

Qy       1309 KTSKKGSKKKSGPPAPEEEEEEERQSGNTE 1338
Db        804 KPINKGRKR----AAVGQESPGGLEAGNAK 829
```

RESULT 8
Q8IXT6
ID   Q8IXT6   PRELIMINARY;   PRT;   529 AA.
AC   Q8IXT6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to androgen-induced prostate proliferative shutoff associated
DE   protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]

```
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Strausberg R.;
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC039256; AAH39256.1; -.
SQ   SEQUENCE   529 AA;  61414 MW;  5C50973ACB920725 CRC64;

  Query Match             35.0%;  Score 2517;  DB 4;  Length 529;
  Best Local Similarity   98.0%;  Pred. No. 1.6e-137;
  Matches  488;  Conservative    4;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLA 60

Qy         61 LHLASDFFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDT 120
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHLASDFFLKHPDKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDT 120

Qy        121 KSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDL 180

Qy        181 MSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTFFNQV 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        181 MSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITNFFNQV 240

Qy        241 LMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMF 300

Qy        301 GAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSH 360
              |||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||
Db        301 GAKDSELASQNKPLWQCYLGRFNDINVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSH 360

Qy        361 DPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 DPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYA 420

Qy        421 LQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYL 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYL 480

Qy        481 YATLDLNAVKALNEMWKC 498
              |||||||||| :: :  |
Db        481 YATLDLNAVKYVSNIKFC 498

RESULT 9
Q96KV4
ID   Q96KV4      PRELIMINARY;      PRT;   479 AA.
AC   Q96KV4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   49J10.1.2 (Androgen-induced prostate proliferative shutoff associated
DE   protein, isoform 2) (Fragment).
GN   49J10.1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Whiteley M.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z84572; CAC94787.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   479 AA;  54103 MW;  AB57E74AAE9656DF CRC64;

  Query Match             34.6%;  Score 2487;  DB 4;  Length 479;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-136;
  Matches  479;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        913 DECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREY 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 DECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREY 60

Qy        973 LKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKN 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKN 120

Qy       1033 ENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPV 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPV 180

Qy       1093 LPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRM 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRM 240

Qy       1153 ETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKRDDSDLVRSELE 1212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKRDDSDLVRSELE 300

Qy       1213 KPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESDEQQWPEEKRLKE 1272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESDEQQWPEEKRLKE 360

Qy       1273 DILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEER 1332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 DILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEER 420

Qy       1333 QSGNTEQKSKSKQHRVSRRAQQRAESPESSAIESTQSTPQKGRGRPSKTPSPSQPKKNV 1391
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 QSGNTEQKSKSKQHRVSRRAQQRAESPESSAIESTQSTPQKGRGRPSKTPSPSQPKKNV 479

RESULT 10
Q96KV3
ID   Q96KV3      PRELIMINARY;      PRT;   536 AA.
AC   Q96KV3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   49J10.1.1 (KIAA0979, isoform 1) (Fragment).
GN   49J10.1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Whiteley M.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z84572; CAC94788.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   536 AA;  60577 MW;  3A30B33F56C81BBC CRC64;

  Query Match             34.4%;  Score 2476.5;  DB 4;  Length 536;
  Best Local Similarity   99.8%;  Pred. No. 3.6e-135;
  Matches  479;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy        913 DECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREY 972
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 DECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREY 60

Qy        973 LKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKN 1032
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKN 120

Qy       1033 ENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPV 1092
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPV 180

Qy       1093 LPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRM 1152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 LPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRM 240

Qy       1153 ETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKRDDSDLVRSELE 1212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKRDDSDLVRSELE 300

Qy       1213 KPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESDEQQWPEEKRLKE 1272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESDEQQWPEEKRLKE 360

Qy       1273 DILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEER 1332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 DILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEER 420

Qy       1333 QSGNTEQKSKSKQHRVSRRAQQ-RAESPESSAIESTQSTPQKGRGRPSKTPSPSQPKKNV 1391
              |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        421 QSGNTEQKSKSKQHRVSRRAQQSRAESPESSAIESTQSTPQKGRGRPSKTPSPSQPKKNV 480
```

RESULT 11
Q8N7J4

```
ID   Q8N7J4      PRELIMINARY;      PRT;   600 AA.
AC   Q8N7J4;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein FLJ41012.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Uterus;
RA   Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA   Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
RA   Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK098331; BAC05286.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   600 AA;  69000 MW;  12917E7912C86BFD CRC64;

  Query Match            34.4%;  Score 2474;  DB 4;  Length 600;
  Best Local Similarity  80.6%;  Pred. No. 5.9e-135;
  Matches  460;  Conservative   64;  Mismatches   47;  Indels    0;  Gaps    0;

Qy         10 DGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLALHLASDFFL 69
              |||| ||||||||:|||: :||::|||||||||||||||||:||: || |||||||:|||
Db         20 DGKIAYPPGVKEITDKITTDEMIKRLKMVVKTFMDMDQDSEDEKQQYLPLALHLASEFFL 79

Qy         70 KHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRYF 129
              ::| |||||||||||||||||||||||||| |||||||:|||||||||||||||||||||
Db         80 RNPNKDVRLLVACCLADIFRIYAPEAPYTSHDKLKDIFLFITRQLKGLEDTKSPQFNRYF 139

Qy        130 YLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSSIICEGD 189
              |||||:|||||||||||||| |||| ||:||||||||| ||:|| |||:||||||| |||
Db        140 YLLENLAWVKSYNICFELEDCNEIFIQLFRTLFSVINNSHNKKVQMHMLDLMSSIIMEGD 199

Qy        190 TVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTFFNQVLMLGKTSIS 249
               |:|||||::|:||:|||||||||::|||| ||||| | || | |  ||||||:||::|:|
Db        200 GVTQELLDSILINLIPAHKNLNKQSFDLAKVLLKRTVQTIEACIANFFNQVLVLGRSSVS 259

Qy        250 DLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMFGAKDSELAS 309
              |||||||||| ||: || |||||||:||||||||||| |||| ||:||||:||:|||:||:
Db        260 DLSEHVFDLIQELFAIDPHLLLSVMPQLEFKLKSNDGEERLAVVRLLAKLFGSKDSDLAT 319

Qy        310 QNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHD 369
              ||:|||||:||||||||||:||| ||||||||||||||||||||||||||||||||||||
Db        320 QNRPLWQCFLGRFNDIHVPVRLESVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHD 379

Qy        370 VIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDA 429
              |||:|:||||:|: |||| || |||||||||||||||||||||||:|||| | |   |||:|
Db        380 VIVTIITAAKRDLALVNDQLLGFVRERTLDKRWRVRKEAMMGLAQLYKKYCLHGEAGKEA 439

Qy        430 AKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAV 489
              |::::|||||||||||||||||:||||:|||||:||||||| |||||||||||:|| |||
Db        440 AEKVSWIKDKLLHIYYQNSIDDKLLVEKIFAQYLVPHNLETEERMKCLYYLYASLDPNAV 499

Qy        490 KALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKF 549
              |||||||||||:||  |::|||| |||  :|: | |:| |:| | :|||||||||||:|||
Db        500 KALNEMWKCQNMLRSHVRELLDLHKQPTSEANCSAMFGKLMTIAKNLPDPGKAQDFVKKF 559

Qy        550 TQVLEDDEKIRKQLEVLVSPTCSCKQAEGCV 580
               ||| ||||:| |||:|:|||||||||: ||
Db        560 NQVLGDDEKLRSQLELLISPTCSCKQADICV 590
```

RESULT 12
Q8VDS0

```
ID   Q8VDS0      PRELIMINARY;      PRT;   584 AA.
AC   Q8VDS0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical 65.8 kDa protein.
GN   9030416H16RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC021408; AAH21408.1; -.
DR   MGD; MGI:1918771; 9030416H16Rik.
KW   Hypothetical protein.
SQ   SEQUENCE   584 AA;  65801 MW;  C7015B0681432851 CRC64;

  Query Match            32.2%;  Score 2317;  DB 11;  Length 584;
  Best Local Similarity  74.3%;  Pred. No. 6.8e-126;
  Matches  434;  Conservative   74;  Mismatches   76;  Indels    0;  Gaps    0;

Qy        525 IFSKVMVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREIT 584
              :| |:| | :|||||||||||:||| ||| ||||:| |||:|:|||||||||: ||||| 
Db          1 MFGKLMTIAKNLPDPGKAQDFVKKFNQVLGDDEKLRSQLELLISPTCSCKQADVCVREIA 60

Qy        585 KKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQ 644
              :|| ||||||||||||:|||||||||||||:|:||||:| :||||:||||||:|||  | 
Db         61 RKLANPKQPTNPFLEMVKFLLERIAPVHIDSEAISALVKLMNKSIEGTADDEEEGVSPDS 120

Qy        645 AIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFP 704
              |||:||||||||||||| |||||||:|||| ||:|:|:||||||:|||:||| ||| |:|
Db        121 AIRSGLELLKVLSFTHPTSFHSAETYESLLQCLRMEDDKVAEAAIQIFRNTGHKIETDLP 180

Qy        705 HIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLI 764
               ||| |:|:||:|:|:| | ||| :||||||||:|| |:||||||| :||:    | ||
Db        181 QIRSTLIPILHQKAKRGTPHQAKQAVHCIHAIFSNKEVQLAQIFEPLSRSLNADVPEQLI 240

Qy        765 TPLVTIGHIALLAPDQFAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMV 824
              ||||::|||::|||||||:| || || |||||||||||  |:|  ||| ||||||||  :
Db        241 TPLVSLGHISMLAPDQFASPMKSVVANFIVKDLLMNDRSTGEKNGKLWSPDEEVSPEVLA 300

Qy        825 KIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGS 884
              |:||||::|||||||||| |||  ||||||: :| |:|||||| :||| |||||||||||
Db        301 KVQAIKLLVRWLLGMKNNQSKSANSTLRLLSAMLVSEGDLTEQKRISKSDMSRLRLAAGS 360

Qy        885 AIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICA 944
              ||:||||||||||||| ||:|||||:||||||||||:||||||| | :| |||||||| |
Db        361 AIMKLAQEPCYHEIITPEQFQLCALVINDECYQVRQIFAQKLHKALVKLLLPLEYMAIFA 420

Qy        945 LCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPD 1004
```

```
                  |||||||||||||||||:|||::||||:||:    :||||||||||||| |||||||||
Db          421 LCAKDPVKERRAHARQCLLKNISIRREYIKQNPMATEKLLSLLPEYVVPYMIHLLAHDPD 480

Qy         1005 YVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEK 1064
                : : ||::||:|:||||||:||:|| ||||||||||::|| |||| |:||| ||::| |||
Db          481 FTRSQDVDQLRDIKECLWFMLEVLMTKNENNSHAFMKKMAENIKLTRDAQSPDESKTNEK 540

Qy         1065 LYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNT 1108
                ||||||||: :| |||   : :|||||||| :||||||:|  | |
Db          541 LYTVCDVALCVINSKSALCNADSPKDPVLPMKFFTQPEKVISLT 584


RESULT 13
Q9V646
ID   Q9V646      PRELIMINARY;      PRT;  1218 AA.
AC   Q9V646;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CG17509 protein.
GN   CG17509.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003823; AAF58593.1; -.
DR   FlyBase; FBgn0033676; CG17509.
SQ   SEQUENCE   1218 AA;  138887 MW;  F4A6B7238A0801C2 CRC64;

  Query Match             27.6%;  Score 1982.5;  DB 5;  Length 1218;
  Best Local Similarity   37.3%;  Pred. No. 4e-106;
  Matches  444;  Conservative  247;  Mismatches  421;  Indels   77;  Gaps   22;

Qy           13 ITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLALHLASDFFLKHP 72
                | || | : : : :   :|::|||| :        ||||     :: |: ||||| |||::||
Db            4 IVYPTGCRPLVEDLGTDELIRRLKTLANVLQTMDQDDNLYQQ-YIPLALHLLDDFFMQHP 62

Qy           73 GKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRYFYLL 132
                 :||:||:|||:||: |:|||||||  |::| || |  :|| ||:| : | | ||||||
Db           63 SRDVQLLIACCVADVLRVYAPEAPYKEQDQIKTIFKFFIKQLHGLKDPRDPSFKRYFYLL 122

Qy          133 ENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSSIICEGDTVS 192
                ||:|:|||:|:||||||  |||  |:  |:|  ::|: |: ||    :|::| :| | | :|
Db          123 ENLAFVKSFNMCFELEDCQEIFQDLFSTIFKIVNDQHSVKVTNFFLDVLSPLITEADNLS 182

Qy          193 QELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTFFNQVLMLGK--TSISD 250
                 |||| :|:|:|   :|: || |   | : || :|  |:| |   |||: |:: |   | :|
Db          183 VELLDLILINIVEPYKSNNKFACQLTEQLLTKTGDALESTIKMFFNRALVMDKPNTKLS- 241

Qy          251 LSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQ 310
                ::   ::|:| || |:: || ||||||| || | |: |||:    |||::||  |||:|| :
Db          242 ITNKIYDIIYELNRINAGLLCSVLPQLENKLLSTDDAERLKATTLLSRMFSEKDSQLAKK 301

Qy          311 NKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDV 370
                   | : : ||| ||  |:|::||: : | |:||| |   |:|| |::|:|| :| :||:|
Db          302 YPNLLKIFFGRFCDITEPVRIKCVQSSMHFLLNHPSLQHDITEKLRLRNHDLDEVVRHEV 361

Qy          371 IVSIVTAAKKDILLVND--HLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKD 428
                :::||  ||:|  ||  :     ||  |||||||::::|::|:|| ||| |||:     :
Db          362 VMAIVETAKRDFTLVLEAPDLLEIVRERTLDKKYKIRRDAMNGLAYIYKRAICEPNDLST 421

Qy          429 AAK-QIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLN 487
                  | :: |||:|:|| ||:  ::||||||:    :||: |    |||| |||:|    ||| |
Db          422 GLKVRVDWIKNKILHGYYKVGLEDRLLVERLLITCLVPYKLAPEERMKKLYHLLGDLDAN 481

Qy          488 AVKALNEMWKCQNLLRHQVKDLLDLIKQPK-TDASVKAIFSKVMVITRNLPDPGKAQDFM 546
                 | ||  |: | |    |: | | : |      : |     :   : :|   | : |||| || :::
Db          482 ATKAFVELQKNQMKTRNTVSDWIKLHHSKEFTPRVLSQLSAKQANIAKLLPDPLKAAEYL 541

Qy          547 KKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLLE 606
                 :|:  |  | :: : : :::      ||::      :  : ||||    | :| :    :| |:|
Db          542 TQFSNNLRKDAQLLRCINIVLKRDVSCRECADTMGVLLKKLGAHVQ-SNLYYNTVKMLIE 600

Qy          607 RIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHS 666
                |:|· | :| |||  ||    : : |:    |:    |:   ||:||  ||:|:   |:      :
Db          601 RVASVMVDKESIGVLISLIEQCIEKGSMCEEIGISAQEAGERGLKLLAMLSYVFSAHFFT 660

Qy          667 AETFESLLACLKMDDEKVAEAALQIFKNTG---SKIEEDFPHIRSALLPVLHHKSKKGPP 723
                  :   |::|   : : ||   |:     : |       |::  | |   | ||    :  | |
Db          661 DTSLRHLISLLSYEQDYVAPLVLKTLTHLGRYQPLIDDPTPAILDELAPVCKDFALIGTP 720

Qy          724 RQAKYAIHCIH---------------AIFSSKETQ-----FAQIFEPLHKSLDPSNLEHL 763
                :|||:|: ||               |  :|  ||      | :| | |   | | | | ||
Db          721 KQAKHAVRCIFVNSQSSASTDGATSGAGSASTTTQTVHPIFNEIIETLRLKLTP-NCEHQ 779

Qy          764 ITPLVTIGHIALLAPDQFAAPWKSWVATFIVKDLLMNDRLPGKKTTKL-----WVPDEEV 818
                  | :||:||||  |   |  | |: :|  |||:||: : :| ::  :|         |  |::
Db          780 RTKIVTLGHIAYNMPQAFLTPIKNMIARRIVKELLIQE-VPAQRDYELPEDSDWCAQEKL 838

Qy          819 SPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRL 878
                 |:|: |:  |:  |:| | |||||:: :   | :|    ::  |||  |::   :|   | |
Db          839 PPDTLCKLDALKAMARWLLGLRTD-EHAAQKTFRMLAAFVNQRGDLLGQNRLCGAEKSWL 897

Qy          879 RLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLR---- 934
                || |  |::|: ::      :   : |||      :  :  |    :||::||:||||||||
Db          898 RLGAACAMLKVCEQKGVGDQYSAEQYLQLSQLMADPVPEVREIFARKLHKGLSRSLPRNC 957

Qy          935 LPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVS------EKLLSLLP 988
                |||::|  :  |   :  : ::  :   |       ::| ||||||  |  |      : | :||
Db          958 LPLDFMGLYVLAGLETERKLQDLVRHYAETDVNKRREYLKTVAMTSPDSSTESQSLHILP 1017

Qy          989 EYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIK 1048
                :|:: : | :| ||| :   :|  ||: :::|| |:|| |||| |    |:| ::::: ||
```

```
Db       1018 DYMLAFAIPVLVHDPRFTNHEDYVQLRKMEKCLRFILEPLMAKRETFVHSFYKQLLQLIK 1077

Qy       1049 QTKDAQGPDDAKMNEKLYTVCDVAMNIIMSK-------STTYSLESPKDPVLPARFFTQP 1101
               : :  |  | : |  |:: :||:|| || ||        ::|:|:       ||  :: :|
Db       1078 HREFSLG-SDKRDNYKMWALCDLAMYIIDSKFSPFDGNTSTFSMPL----ALPEMYYKEP 1132

Qy       1102 D-KNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKS 1149
               || |    |:| :: :               |||  |  |  |   : | |
Db       1133 AVANFQNNDVYIPLDVYT-----------LGA--KSTSKAAATAMTTS 1167


RESULT 14
Q9HFF5
ID   Q9HFF5      PRELIMINARY;      PRT;  1205 AA.
AC   Q9HFF5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Protein with possible role in chromosome dynamics by similarity to
DE   yeast pds5.
GN   SPAC110.02 OR PDS5+.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972h-;
RA   Zimmermann W., Wambutt R., Wood V., Rajandream M.A., Barrell B.G.;
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=L972;
RX   MEDLINE=21481773; PubMed=11598020;
RA   Tanaka K., Hao Z., Kai M., Okayama H.;
RT   "Establishment and maintenance of sister chromatid cohesion in fission
RT   yeast by a unique mechanism.";
RL   EMBO J. 20:5779-5790(2001).
DR   EMBL; AL441624; CAC08560.1; -.
DR   EMBL; AB067651; BAB71784.1; -.
DR   GeneDB_SPombe; SPAC110.02; -.
SQ   SEQUENCE   1205 AA;  138874 MW;  935DABDE0A5E30FF CRC64;

  Query Match             11.6%;  Score 833.5;  DB 3;  Length 1205;
  Best Local Similarity   23.3%;  Pred. No. 1.1e-39;
  Matches  296;  Conservative  264;  Mismatches  564;  Indels  149;  Gaps   38;

Qy         26 ISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLALHLASDFFLKHPGKDVRLLVACCLA 85
              ::  |:::||: ::      : ||:  :::  :  |:|  | ::   |  |  | :|  ||:
Db         18 LTTSEILKRLRDLLGELTSLSQDT-IDRDSVLPVARSLVNNNLLHHKDKGIRSYTLCCIV 76

Qy         86 DIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSYNICF 145
              :: |: ||:||:|    :|:|||  | : | || : :|   : : : :||::: ||| :
Db         77 ELLRLCAPDAPFTL-SQLEDIFQVILKILSGLMNQESTYYPQIYEILESLSNVKSAVLIV 135

Qy        146 ELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSSIICEGDTVSQELLDTVLVNLV- 204
              :| :: |      ::|  | :    |  : |  :|:|::: :| | :|:     |: :   |:
Db        136 DLPNAEEFLVNIFRLFFDLARKGTTKNVEFYMLDIINQLINEINTIPAAALNILFAQLIS 195

Qy        205 ---------PAHKNLNKQAYDLAKALLKRTAQAIEPYITTFFNQVLMLGKTSISD---LS 252
                       :     :  |: ||: :    :|   :: |:  :|: ::    : |:||
Db        196 GKGVRQTIGSSDSTNHGPAFQLARNIFHDSADRLQRYVCQYFSDIIFDSRDSLSDSMTTP 255

Qy        253 EHVF--DLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQ 310
               | :|  :|:|:|:            ||::::|| | :|::        ||  :  :  |   :|: : |
Db        256 EFIFSHNLVLQLWKYAPTTLLNIIPQFENELQAEQTSVRLVAIETVGLML--QDNAIWSD 313

Qy        311 NKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNH---PDLAKDLTEYLKVRSHDPEEAIR 367
                :|  : || ||  |  |::|:: ||: | |        :: ::: : |: :  : | : :|  :|
Db        314 YPRVWSAFCGRLNDKSVACRIKCIEVASNALQNSLATSEIIENVVQMLQSKLADTDEKVR 373

Qy        368 HDVIVSIVTAAKKDI-LLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAG 426
                :  :|     :   :   :  |  :  :|  |::  || :|:  |:|||  :          |
Db        374 VATLKTIEQLTFETFKMQFSVQALKLMGDRLRDRKLNVRLQAIRTLSQIYNRAYQDLIDG 433

Qy        427 KDAAKQI-AWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPH-NLETTERMKCLYYLYATL 484
                : : |: :||   || ::| |         ||    |: : : : :  :|  |:  |:
Db        434 VEYSIQMFSWIPSSLLEVFYVNDETTNAAVEICMAELVLQYLSSDTQTRLNRLFLSIKYF 493

Qy        485 DLNAVKA----LNEMWKCQNLLRHQVK-------DLLDLIKQPKTDASVKAIFSKVMVIT 533
                |::         |      |  ||  : ::          ::|   ::  |:   |  |       :|:
Db        494 SEKAMRVFILLLQRQVKYSELLNYYIECCKNYNGGVMDNDEESITNKLKKVI----DIIS 549

Qy        534 RNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQP 593
                  :|    :    :||  : |  |  :  :  |:       |  |  |          |
Db        550 SKSSNPTLTEATFRKFAE-LNDRQSYKMLLQ-----TFSIKSEYQVV------------ 590

Qy        594 TNPFLEMIKFLLERIAPVHIDTESI---------SALIKQVNKSIDGTADDEDGVPTDQ 644
                 |: ||:| :|:: :  | |:               |||        ::           :| |
Db        591 ----LKSIKYLFKRVSET-LSTASLECFRIFVYRSALFAFNKSNVHEIIQLLNEPVKYHN 645

Qy        645 AIRAGLELLKVLSFTHPISFHSA--ETFESLLACLKMDDEKVAEAALQIFKNTGSKIEED 702
               ::      ||: |      ||    :           |     :::          | || :|  |          || :::
Db        646 FLKPSEALLQHLPLIHPNIYGEVVIEVENIIVSSGIESDPKVIKALSQF-----SKRKKN 700

Qy        703 FPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEH 762
              |  |::     :|      |         |||  |   ||  :||  :         :          |||:
Db        701 F-SIQTTTAEILRKLCLHGTQEQAKQAATII-AITETKEFKLDMITNIV------ENLEY 752

Qy        763 ---LITPLVTIGHIALLAPDQFAAPWKSWVATFIVKDLLMNDRLPGK----KTTKLWVPD 815
                |    |:|:|  :  |     ::        |   |:||  ::   | | |         : |
Db        753 NGGLPVRLMTLGQLFLYTLEE-VEKVADQVTEFLVKKVI--QRFPEKYDDTHNDEEWCTY 809

Qy        816 EEVSPETMVKIQAIKMMVRWLLGMKNNHS--KSGTSTLRLLTTILHSDGDLTEQGKISKP 873
              |::   || |: ||:::|  |                  |     ::||  :|  :|:|:          |
Db        810 EKLDNLTMCKVLAIRVLVNRLRAAAGGTEALNIGAPIIKLLKVLLMADGELSPFKNTPKI 869

Qy        874 DMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRL 933
                : |||  |        :||   |  |  |  |      |  :|   || : ||  :|  ||  |
Db        870 SRAYLRLTASKYFLKLCSIPFYAEHIDFSSYVQISLLCQDENFDVRNLFLTKLQKQLQLK 929

Qy        934 RLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVP 993
               :||: |   :   | | || :| :  |    |                 :::       |  :       |||
Db        930 KLPISYYPLLFLTAVDPEEEIKTKA-----------SIWIRSQVAFFQKTHDFTMEYVAT 978

Qy        994 YTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDA 1053
               | ||||:| ||     ::       |    :     ::||   :: ::|          :   :::  |||: |
Db        979 YLIHLLSHHPDISSIESENSLDFIAYIRFYVDTVVNSENV----PIVFHLMQRIKQSYDV 1034

Qy       1054 QGPDDAKMNEKLYTVCDVAMNIIMSKS-------TTY--SLESPKDPVLPARFFTQPDKN 1104
               :|   |  :| : |:| |: ||           |||    :: | : : |              :  :
Db       1035 --IEDG--NNYIYVLSDMAQKILQVKSQNFGWSLTTYPKQIKLPYEILRPIPSIDEKKRI 1090

Qy       1105 FSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNP 1164
              |:   | :: |:|:|                    :  |:   |:||   ||:  | : ::          || |:
Db       1091 FN--KIFITPKMES---------QIEHAIRTPVSSFAKQTTNKHANLKQKKTHSSKSDK 1138

Qy       1165 SSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTE 1224
               |  |  |            |    ||           :| : |  :        :  ||              ::|
Db       1139 KSSRRRK-----NEKRRKLNEQNPNIRNVPERSSSRFQGIRINYSEAPS----SSEEISE 1189

Qy       1225 QEEKLGMDDLTKL 1237
               :||::  :|  ::
Db       1190 EEEEISEEDFDEI 1202


RESULT 15
Q9FIL0
ID   Q9FIL0      PRELIMINARY;      PRT;  1638 AA.
AC   Q9FIL0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Gb|AAF07790.1.
OS   Arabidopsis thaliana (Mouse-ear cress).
```

```
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Columbia;
RX   MEDLINE=99156233; PubMed=10048488;
RA   Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.
RT   Sequence features of the regions of 1,081,958 bp covered by seventeen
RT   physically assigned P1 and TAC clones.";
RL   DNA Res. 5:379-391(1998).
DR   EMBL; AB016886; BAB11316.1; -.
DR   EMBL; AB025628; BAB11316.1; JOINED.
DR   EMBL; AB028612; BAB11316.1; JOINED.
DR   InterPro; IPR001395; Aldo/ket_red.
DR   PROSITE; PS00063; ALDOKETO_REDUCTASE_3; 1.
SQ   SEQUENCE   1638 AA;  185115 MW;  0F5BB6ABE1888F93 CRC64;

  Query Match             11.2%;  Score 808;  DB 10;  Length 1638;
  Best Local Similarity   23.0%;  Pred. No. 5e-38;
  Matches  337;  Conservative  278;  Mismatches  634;  Indels  216;  Gaps   48;

Qy         19 VKEISDK-----ISKEEMVRRLKMVVKTFMDMDQDSE----EEKELYLNLALHLASDFFL 69
              :||:  |     :||: ::: ||       :::|       :  : :|:  :      |
Db          9 LKELGSKLDLAPVSKDSLLKLLKEAAVCLSELEQSPPPAVLKSIQPFLDAVI---KPEIL 65

Qy         70 KHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRYF 129
                |  |||:||||  |:::| || |||||| | : :||||  |       || |    | | |
Db         66 NHQDKDVKLLVASCVSEITRITAPEAPY-SDNIMKDIFQLIVSAFAGLNDVSGPSFGRRV 124

Qy        130 YLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSSIICEGD 189
              :|| :|  :|   :  :||  :::  :::| |     |  :  |  |    | ::|  :: | :
Db        125 LILETVAKYRSCVVMLDLE-CDDLVKEVFTTFLDVARDDHPEIVFSSMQNIMIVLLEESE 183

Qy        190 TVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTFFNQVLMLGKTSIS 249
               | : || :| |       ::    |  ||  :::   |  :|  |    |       | | : |
Db        184 DVQEHLLLILLSKLGRNRSDVRDAARRLAMKVIEHCAPKVESDIKQFLIS-SMSGDSRFS 242

Qy        250 DLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMFGAKDSELAS 309
                    ::| :||          |   |  |  |   :|  ::   | ||:||  |: ::|          ::
Db        243 SSQIDYHEVIYDLYRCAPQALSGVAPYLTGELLADKLETRLKVVGLVGELFSLPGRVISE 302

Qy        310 QNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHP--DLAKDLTEYLKVRSHDPEEAIR 367
              :    ::  :| |   |   | |:|:   :          ||::  |        |     :    |   |  |  :| ||
Db        303 EFDSIFLEFLKRLTDRVVEVRMAILDHIKDCLLSDPLRAEASQIISALCDRLLDYDENIR 362

Qy        368 HDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGK 427
                |:   |     :    :     :      :   | ||   ||    |:      |   |  :::: |  |: |  ||
Db        363 KQVVAVICDVSVSALTSIPVDTMKLVAERLRDKAILVKTYTMERLTELFRVYCLRCADGK 422

Qy        428 DAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLN 487
                    ||  |:|   |               :| |        : |  :      :::|        ::: |
Db        423 VDTGDFNWIPGKILRCLYDKDFRSD-TIEYILCSSLFPSDFSVRDKVKHWIQIFSGFDKV 481

Qy        488 AVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVM----VITRNLPDPGKA- 542
                  ||  :: :  :   :: :::   |  |||  :    |     |    |::     |::|    ||  |
Db        482 ETKAFEKILEQRQRIQQEMQRYLS-IKQTQQTADAPEIQKKILFGFRVMSRAFSDPPKTE 540

Qy        543 QDFMKKFTQVLED--DEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEM 600
              |:|:      :|:    |    |  |  |  |   |:  |    |    ||     :|:     |:  :  |     ||
Db        541 QNFL-----ILDQLKDANIWKILTNLLDPNTSITQASR-IRDDMLKILSEKHSLYDFLST 594

Qy        601 IKFLLERIAPVHIDTESISALIKQVN--KSIDGTADDEDEGVPTDQAIRAGLELLKVLSF 658
              :     : : :          |  :  :  :|:   ||   |              |:  :  :   |   :|:
Db        595 LSI---KCSYLLFSKEYVKEILAEVSVRKSSKNTL-----------GIQPCMDFLGLLAC 640

Qy        659 THPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKS 718
                |   |   ||   |  |::  || |||  :  |   |:|        |    |   |:       :  |::   :|
Db        641 FCPSLFDGAE--EELISFLKDDDEMMKEGTLKILAKAGGTIRENLIVLASSVDLLLERIC 698

Qy        719 KKGPPRQAKYAIHCIHAI-----FSSKETQFAQIFEPLH-KSLDPSNLEHLITPLVTIGH 772
               :|   :|||||:|  :  :|            |        : :: :  |   |     |:  |:              :|
Db        699 VEGNRKQAKYAVHALASITKDDGLKSLSVLYKRLVDMLEDKRYQPAVLQ-------CLGC 751

Qy        773 IALLAPDQFAAPWKSWVATFIVKDL--LMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIK 830
              || :|      :       :| |   ||     :     |  | : :      || :      |         |:           :||   ||
Db        752 IAQIAMPVYETR-ESEVVEFIRSKILKLKSETVDDKKLS--WDDKSEI---CQLKIYGIK 805

Qy        831 MMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLA 890
               :|:   |    |:        ::|       |    :         :    |  |:::|     :     |    |  :  |||||   |:::|:
Db        806 TLVKSYLPFKDAQLRAGVDDLLGILKNILSFGEVSEDLESSSVDKAHLRLAAAKAVLRLS 865

Qy        891 QEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALC---- 946
              :      :  :   |  :|  :   |                  :::|   |:|: :        |  ::|     |:
Db        866 RH--WDDKIPIEIFHLTLKTPEIPFPTAKKIFLGKVHQYVKDRVLEMKY--ACSFLFDIT 921

Qy        947 ---------AKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIH 997
                       |   :  :        |:  |     |:  |:                 |               ::|  |  :::||  :|
Db        922 GSNVLESEEDKHNLADIIQHSYQTKVRKIS---------AQTDANSVTLYPHHILPYLVH 972

Qy        998 LLAHD--PDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRK---------MVE 1045
                |||      ||    |   : |::: :   :         |:  ::  : |:   |  |:                  ||                :
Db        973 ALAHHSCPDVEKCKDVKEYEMIYRQLYLIISMLLHKEEDGKTEDIDKEREYVPTIILIFH 1032

Qy       1046 NIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQ--PDK 1103
              :|||::|        ||    ::       :  :|::  ::||                                   ||   ||
Db       1033 SIKQSEDV---TDATKSKNSHAICELGLSII------------------NHLTQKEPDL 1070

Qy       1104 NFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRM----ETV---- 1155
                    |       |||        :  :   |  :                            |   :||        ::          |||
Db       1071 QGEITPVSLPP---TLYKPSE-------------KVEGDKSQVGEEKLWLADETVLLHF 1113

Qy       1156 ------SNASSSSNPSSPGR---IKGRLDSSEMDHSENEDYTMSSPLPGKKSDKRDDSDL 1206
                    |:|  :|    |  :                  |   |     |   :|:                  ||   ||      ::
Db       1114 RALKLESHADASVIPQTSENEVMIDGESDGNEI------------PL-GKIVER------ 1154

Qy       1207 VRSELEKPRGRKKTPVTEQEEKLGMD--DLTKLVQE------QKPKGSQRSRKRGHTASE 1258
              :|::   |   |   ||              |::  |   :    |:  |:|:|             |              :    |             |:  ||
Db       1155 LRAQGTKTRKGKKNKSVPAEDENGKNDVDVLKMVREINLDHLQMLDKFESSNGHKHSPSE 1214

Qy       1259 SDEQQWPEEKRLKEDILENEDEQNSPP-KKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKK 1317
                   |        ::|      |  : :        :   |     ::         |     |                ||   :                        |
Db       1215 RAEICQRDQKGNKRNVGDATSVVSVPKRRRSSSGHSPYKFSNSGPKVQ----------LK 1264

Qy       1318 KSGPPAPEEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESPESSAIESTQS-------- 1369
                   |          |   :  : ::    |  ::    ::    ::    ::    :      :  :: :       ||  ::  |:|
Db       1265 ASEDELHLESDMDKNVSLDSHDENSDQEKMLESISPRKRKKSLSSKLKITESDWALTDVE 1324

Qy       1370 TPQKGRGRPSKTPSPS---QPKKNV 1391
                  :     |    ||      |    |        :  :|||
Db       1325 RSRSAGGGDSKLKSASGSMKKRKNV 1349

Search completed: September 24, 2003, 20:20:43
Job time : 180 secs
```

NPA


```
OM protein - protein search, using sw model

Run on:         September 24, 2003, 16:31:58 ; Search time 20 Seconds
                                                (without alignments)
                                                3270.707 Million cell updates/sec

Title:          US-09-512-581B-2
Perfect score:  7193
Sequence:       1 MAHSKTRTNDGKITYPPGVK..........QKGRGRPSKTPSPSQPKKNV 1391

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters:      127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_41:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	833.5	11.6	1205	1	PDS5_SCHPO	Q9hff5 schizosacch
2	557	7.7	1277	1	PDS5_YEAST	Q04264 saccharomyc
3	239.5	3.3	5171	1	BPEA_HUMAN	O94833 homo sapien
4	235.5	3.3	1875	1	MLP1_YEAST	Q02455 saccharomyc
5	229	3.2	2663	1	CENE_HUMAN	Q02224 homo sapien
6	227	3.2	1916	1	RIF1_YEAST	P29539 saccharomyc
7	220.5	3.1	1727	1	ALM1_SCHPO	Q9utk5 schizosacch
8	217.5	3.0	1020	1	NFH_HUMAN	P12036 homo sapien
9	217	3.0	1790	1	USO1_YEAST	P25386 saccharomyc
10	212.5	3.0	8545	1	ANC1_CAEEL	Q9n4m4 caenorhabdi
11	212	2.9	831	1	NFH_RAT	P16884 rattus norv
12	210.5	2.9	5596	1	MDN1_HUMAN	Q9nu22 homo sapien
13	209	2.9	7389	1	BPA1_MOUSE	Q91zu6 mus musculu
14	208	2.9	1153	1	A3D1_HUMAN	O14617 homo sapien
15	206.5	2.9	699	1	NP14_HUMAN	Q14978 homo sapien
16	206	2.9	6632	1	UN89_CAEEL	O01761 caenorhabdi
17	205.5	2.9	2230	1	GOG4_HUMAN	Q13439 homo sapien
18	205	2.8	2349	1	TPR_HUMAN	P12270 homo sapien
19	203.5	2.8	2459	1	MAPB_RAT	P15205 rattus norv
20	202	2.8	2748	1	NUM1_YEAST	Q00402 saccharomyc
21	201	2.8	700	1	TRDN_CANFA	P82179 canis famil
22	199	2.8	1359	1	ATRX_CAEEL	Q9u7e0 caenorhabdi
23	199	2.8	2464	1	MAPB_MOUSE	P14873 mus musculu
24	197.5	2.7	3259	1	GIAN_HUMAN	Q14789 homo sapien
25	195.5	2.7	1679	1	YIO9_YEAST	P40457 saccharomyc
26	195	2.7	488	1	CYL2_BOVIN	Q28092 bos taurus
27	194	2.7	705	1	TRDN_RABIT	Q28820 oryctolagus
28	194	2.7	1533	1	TP2A_PIG	O46374 sus scrofa
29	194	2.7	1539	1	Y373_HUMAN	O15078 homo sapien
30	193.5	2.7	3210	1	CENF_HUMAN	P49454 homo sapien
31	192.5	2.7	2469	1	TEGU_HSVSA	Q01056 herpesvirus
32	190.5	2.6	2468	1	MAPB_HUMAN	P46821 homo sapien
33	188.5	2.6	633	1	MLH_TETTH	P40631 tetrahymena
34	188.5	2.6	1453	1	Y373_BOVIN	Q9tu23 bos taurus
35	188	2.6	1841	1	CC12_SCHPO	Q10059 schizosacch
36	187.5	2.6	2017	1	MYSN_DROME	Q99323 drosophila
37	187	2.6	1087	1	NFH_MOUSE	P19246 mus musculu
38	187	2.6	1311	1	ATRX_DROME	Q9gqn5 drosophila
39	186.5	2.6	1130	1	YL17_CAEEL	Q11102 caenorhabdi
40	186	2.6	728	1	TRDN_HUMAN	Q13061 homo sapien
41	186	2.6	2869	1	RBP1_PLAVB	Q00798 plasmodium
42	185.5	2.6	2364	1	SPCO_HUMAN	Q01082 homo sapien
43	184.5	2.6	3678	1	DMD_MOUSE	P11531 mus musculu
44	184.5	2.6	3924	1	ANK2_HUMAN	Q01484 homo sapien
45	184	2.6	1466	1	SPA2_YEAST	P23201 saccharomyc

ALIGNMENTS

```
RESULT 1
PDS5_SCHPO
ID   PDS5_SCHPO     STANDARD;      PRT;  1205 AA.
AC   Q9HFF5; O94237;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Sister chromatid cohesion protein pds5 (Precocious dissociation of
DE   sisters protein 5).
GN   PDS5 OR SPAC110.02.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RC   STRAIN=972;
RX   MEDLINE=21481773; PubMed=11598020;
RA   Tanaka K., Hao Z., Kai M., Okayama H.;
RT   "Establishment and maintenance of sister chromatid cohesion in fission
RT   yeast by a unique mechanism.";
RL   EMBO J. 20:5779-5790(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
RN   [3]
RP   SEQUENCE OF 816-1205 FROM N.A.
RA   Lee M., Yoo H.S., Chung K.S.;
```

```
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   FUNCTION.
RX   MEDLINE=21850851; PubMed=11861765;
RA   Wang S.-W., Read R.L., Norbury C.J.;
RT   "Fission yeast Pds5 is required for accurate chromosome segregation
RT   and for survival after DNA damage or metaphase arrest.";
RL   J. Cell Sci. 115:587-598(2002).
CC   -!- FUNCTION: Required for the establishment and maintenance of sister
CC       chromatid cohesion during S phase. Prevents their formation until
CC       esol is present. May also have a role during meiosis.
CC   -!- SUBUNIT: Interacts with esol.
CC   -!- SUBCELLULAR LOCATION: Nuclear; centromeric. Localized to chromatin
CC       throughout the cell cycle.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB067651; BAB71784.1; -.
DR   EMBL; AL441624; CAC08560.1; -.
DR   EMBL; AF049529; AAD02493.1; -.
DR   GeneDB_SPombe; SPAC110.02; -.
KW   Mitosis; Nuclear protein; Centromere.
FT   CONFLICT    816    818       TMC -> GTS (IN REF. 3).
FT   CONFLICT    902    908       QISLLCQ -> SNLLIMP (IN REF. 3).
SQ   SEQUENCE   1205 AA;  138874 MW;  935DABDE0A5E30FF CRC64;

  Query Match             11.6%;  Score 833.5;  DB 1;  Length 1205;
  Best Local Similarity   23.3%;  Pred. No. 5.2e-31;
  Matches  296;  Conservative  264;  Mismatches  564;  Indels  149;  Gaps   38;

Qy         26 ISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLALHLASDFFLKHPGKDVRLLVACCLA 85
Db         18 LTTSEILKRLRDLLGELTSLSQDT-IDRDSVLPVARSLVNNNLLHHKDKGIRSYTLCCIV 76

Qy         86 DIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSYNICF 145
Db         77 ELLRLCAPDAPFTL-SQLEDIFQVILKILSGLMNQESTYYPQIYEILESLSNVKSAVLIV 135

Qy        146 ELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSSIICEGDTVSQELLDTVLVNLV- 204
Db        136 DLPNAEEFLVNIFRLFFDLARKGTTKNVEFYMLDIINQLINEINTIPAAALNILFAQLIS 195

Qy        205 ---------PAHKNLNKQAYDLAKALLKRTAQAIEPYITTFFNQVLMLGKTSISD---LS 252
Db        196 GKGVRQTIGSSDSTNHGPAFQLARNIFHDSADRLQRYVCQYFSDIIFDSRDSLSDSMTTP 255

Qy        253 EHVF--DLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQ 310
Db        256 EFIFSHNLVLQLWKYAPTTLLNIIPQFENELQAEQTSVRLVAIETVGLML--QDNAIWSD 313

Qy        311 NKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNH---PDLAKDLTEYLKVRSHDPEEAIR 367
Db        314 YPRVWSAFCGRLNDKSVACRIKCIEVASNALQNSLATSEIIENVVQMLQSKLADTDEKVR 373

Qy        368 HDVIVSIVTAAKKDI-LLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAG 426
Db        374 VATLKTIEQLTFETFKMQFSVQALKLMGDRLRDRKLNVRLQAIRTLSQIYNRAYQDLIDG 433

Qy        427 KDAAKQI-AWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPH-NLETTERMKCLYYLYATL 484
Db        434 VEYSIQMFSWIPSSLLEVFYVNDETTNAAVEICMAELVLQYLSSDTQTRLNRLFLSIKYF 493

Qy        485 DLNAVKA----LNEMWKCQNLLRHQVK-------DLLDLIKQPKTDASVKAIFSKVMVIT 533
Db        494 SEKAMRVFILLLQRQVKYSELLNYYIECCKNYNGGVMDNDEESITNKLKKVI----DIIS 549

Qy        534 RNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQP 593
Db        550 SKSSNPTLTEATFRKFAE-LNDRQSYKMLLQ-----TFSIKSEYQVV------------ 590

Qy        594 TNPFLEMIKFLLERIAPVHIDTESI---------SALIKQVNKSIDGTADDEDEGVPTDQ 644
Db        591 ----LKSIKYLFKRVSET-LSTASLECFRIFVYRSALFAFNKSNVHEIIQLLNEPVKYHN 645

Qy        645 AIRAGLELLKVLSFTHPISFHSA--ETFESLLACLKMDDEKVAEAALQIFKNTGSKIEED 702
Db        646 FLKPSEALLQHLPLIHPNIYGEVVIEVENIIVSSGIESDPKVIKALSQF-----SKRKKN 700

Qy        703 FPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEH 762
Db        701 F-SIQTTTAEILRKLCLHGTQEQAKQAATII-AITETKEFKLDMITNIV------ENLEY 752

Qy        763 ---LITPLVTIGHIALLAPDQFAAPWKSWVATFIVKDLLMNDRLPGK----KTTKLWVPD 815
Db        753 NGGLPVRLMTLGQLFLYTLEE-VEKVADQVTEFLVKKVI--QRFPEKYDDTHNDEEWCTY 809

Qy        816 EEVSPETMVKIQAIKMMVRWLLGMKNNHS--KSGTSTLRLLTTILHSDGDLTEQGKISKP 873
Db        810 EKLDNLTMCKVLAIRVLVNRLRAAAGGTEALNIGAPIIKLLKVLLMADGELSPFKNTPKI 869

Qy        874 DMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRL 933
Db        870 SRAYLRLTASKYFLKLCSIPFYAEHIDFSSYVQISLLCQDENFDVRNLFLTKLQKQLQLK 929

Qy        934 RLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVP 993
Db        930 KLPISYYPLLFLTAVDPEEEIKTKA-----------SIWIRSQVAFFQKTHDFTMEYVAT 978

Qy        994 YTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDA 1053
Db        979 YLIHLLSHHPDISSIESENSLDFIAYIRFYVDTVVNSENV----PIVFHLMQRIKQSYDV 1034

Qy       1054 QGPDDAKMNEKLYTVCDVAMNIIMSKS-------TTY--SLESPKDPVLPARFFTQPDKN 1104
Db       1035 --IEDG--NNYIYVLSDMAQKILQVKSQNFGWSLTTYPKQIKLPYEILRPIPSIDEKKRI 1090

Qy       1105 FSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNP 1164
Db       1091 FN--KIFITPKMES----------QIEHAIRTPVSSFAKQTTNKHANLKQKKTHSSKSDK 1138

Qy       1165 SSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTE 1224
Db       1139 KSSRRRK-----NEKRRKLNEQNPNIRNVPERSSSRFQGIRINYSEAPS----SSEEISE 1189

Qy       1225 QEEKLGMDDLTKL 1237
Db       1190 EEEEISEEDFDEI 1202


RESULT 2
PDS5_YEAST
ID   PDS5_YEAST     STANDARD;      PRT;  1277 AA.
AC   Q04264; Q04780;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Sister chromatid cohesion protein PDS5 (Precocious dissociation of
DE   sisters protein 5).
GN   PDS5 OR YMR076C OR YM9582.01C OR YM9916.15C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   PubMed=9169872;
RA   Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA   Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA   Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
```

```
RA   Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   XIII.";
RL   Nature 387:90-93(1997).
RN   [2]
RP   FUNCTION, AND SUBCELLULAR LOCATION.
RX   MEDLINE=20517447; PubMed=11062262;
RA   Hartman T., Stead K., Koshland D., Guacci V.;
RT   "Pds5p is an essential chromosomal protein required for both sister
RT   chromatid cohesion and condensation in Saccharomyces cerevisiae.";
RL   J. Cell Biol. 151:613-626(2000).
RN   [3]
RP   FUNCTION.
RX   MEDLINE=21065936; PubMed=11137006;
RA   Panizza S., Tanaka T., Hochwagen A., Eisenhaber F., Nasmyth K.;
RT   "Pds5 cooperates with cohesin in maintaining sister chromatid
RT   cohesion.";
RL   Curr. Biol. 10:1557-1564(2000).
CC   -!- FUNCTION: Essential for the establishment and maintenance of
CC       sister chromatid cohesion at centromere proximal and distal
CC       regions during S phase. Also required for chromosomal
CC       condensation.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z49259; CAA89222.1; -.
DR   EMBL; Z48952; CAA88801.1; -.
DR   PIR; S54451; S54451.
DR   SGD; S0004681; PDS5.
DR   GO; GO:0005708; C:mitotic chromosome; IDA.
DR   GO; GO:0007076; P:mitotic chromosome condensation; IMP.
DR   GO; GO:0007064; P:mitotic sister chromatid cohesion; IMP.
KW   Mitosis; Nuclear protein.
SQ   SEQUENCE   1277 AA;  147040 MW;  9DF40A5274FD9623 CRC64;

  Query Match             7.7%;  Score 557;  DB 1;  Length 1277;
  Best Local Similarity   20.3%;  Pred. No. 2.7e-18;
  Matches  272;  Conservative  267;  Mismatches  611;  Indels  190;  Gaps   46;

Qy         12 KITYPPGVKEISDK-ISKEEMVRRLKMVVKTFMDMDQDSEEEKEL--YLNLALHLASDFF 68
Db          8 KLKFNSPIISTSDQLISTNELLDRLKALHEELASLDQDNTDLTGLDKYRDA---LVSRKL 64

Qy         69 LKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRY 128
Db         65 LKHKDVGIRAFTACCLSDILRLYAPDAPYTDA-QLTDIFKLVLSQFEQLGDQENGYHIQQ 123

Qy        129 FYLLENIAWVKSYNICFELEDSNEIFTQLYR-----------TLFSVINNGHNQKVHMHM 177
Db        124 TYLITKLLEYRSIVLLADLPSSNNLLIELFHIFYDPNKSFPARLFNVIGG---------- 173

Qy        178 VDLMSSIICEGDTVSQELLDTVL-------VNLVPAHKNLNKQ-AYDLAKALLKRTAQAI 229
Db        174 --ILGEVISEFDSVPLEVLRLIFNKFLTYNPNEIPEGLNVTSDCGYEVSLILCDTYSNRM 231

Qy        230 EPYITTFFNQVLML-----GKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSN 284
Db        232 SRHLTKYYSEIIHEATNDDNNSRLLTVVVKLHKLVLRLWETVPELINAVIGFIYHELSSE 291

Qy        285 DNEERLQVVKLLAKMFGA-KDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMN 343
Db        292 NELFRKEATKLIGQILTSYSDLNFVSTHSDTFKAWISKIADISPDVRVEWTESIPQIIAT 351

Qy        344 HPDLAKDLTEYLKVRSHDPEEAIRHDVIVSI----VTAAKKDIL--LVNDHLLNFVRERT 397
Db        352 REDISKELNQALAKTFIDSDPRVRRTSVMIFNKVPVTEIWKNITNKAIYTSLLHLARE-- 409

Qy        398 LDKRWRVRKEAMMGLAQIYKKYALQ---SAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLL 454
Db        410 --KHKEVRELCINTMAKFYSNSLNEIERTYQNKEIWEIIDTIPSTLYNLYYINDLNINEQ 467

Qy        455 VERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIK 514
Db        468 VDSVIFEYLLPFEPDNDKRVHRLLTVLSHFDKKAFTSFFAFNARQIKISFAISKYIDFSK 527

Qy        515 QPKTDASVKAIFSKVMV---------ITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEV 565
Db        528 FLNNQESMSSSQGPIVMNKYNQTLQWLASGLSDSTKAIDALETIKQF--NDERIFYLLNA 585

Qy        566 LVSPTCSCKQAEGCVREITKKLGNPK------------QPTNPFLEMIKFLLERIAPVH 612
Db        586 CVTNDIPFLTFKNCYNELVSKLQTPGLFKKYNISTGASIMPRD-IAKVIQILLFRASPII 644

Qy        613 IDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLE--LLKVLSFTHPISF-HSAET 669
Db        645 YNVSNISVLLNLSNNS--------------DAKQLDLKRRILDDISKVNPTLFKDQIRT 689

Qy        670 FESLLACL------KMDDEKVAEAALQIFKNTGS-KIEEDFPHIRSALLPVLHHKSKKGP 722
Db        690 LKTIIKDLDDPDAEKNDNLSLEEALKTLYKASKTLKDQVDFDD--TFFFTKLYDFAVESK 747

Qy        723 PRQAKYAIHCIHAIFSSKETQFAQI---FEPLHKSLDPSNLEHLITPLVTIGHIALLAPD 779
Db        748 PEITKYATKLI-ALSPKAEETLKKIKIRILPLDLQKDKYFTSHIIVLMEIFKKFPHVLND 806

Qy        780 QFAAPWKSWVATFIVKDLLMNDRLPGKKTTKL-WVPDEEVSPETMV----KIQAIKMMVR 834
Db        807 D-----STDIISYLIKEVLLSNQVVGDSKKEIDWVEDSLLSDTKYSAIGNKVFTLKLFTN 861

Qy        835 WLLGM-----KNNHSKSGT-STLRLLTTILHSDGDL-TEQGKISKPDMS----RLRLAAG 883
Db        862 KLRSIAPDVPRDELAESFTEKTMKLFFYLIASGGELISEFNKEFYPTPSNYQTKLRCVAG 921

Qy        884 SAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAIC 943
Db        922 IQVLKLARISNLNNFIKPSDIIKLINLVEDESLPVRKTFLEQLKDYVANELISIKFLPLV 981

Qy        944 ALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDP 1003
Db        982 FFTAYEPDVELKTTTKIWI--NFTFGLKSFKK---------GTIFERALPRLIHAIAHHP 1030

Qy       1004 DYVKVQDIE------QLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIK------QTK 1051
Db       1031 DIVGGLDSEGDAYLNALTTAIDYLLFYFDSIAAQENFSLLYYLSERVKNYQDKLVEDEID 1090

Qy       1052 DAQGPDDAK-------MNEKLYTVCDVAMNIIMS----KSTTYS-----LESPKDPVLPA 1095
Db       1091 EEEGPQKEEAPKKHRPYGQKMYIIGELSQMILLNLKEKKNWQHSAYPGKLNLPSDLFKP- 1149

Qy       1096 RFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETV 1155
Db       1150 -FATVQEAQLS-FKTYIPESLT------EKIQNNIKAKIGRILHTSQTQRQRLQKRLLAH 1201

Qy       1156 SNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSD-KRDDSDLVRSELEKP 1214
Db       1202 ENNESQKKKKKVHHARSQADDEEGDGDRESDSDDDSYSPSNKNETKKGHENIV---MKKL 1258

Qy       1215 RGRKKTPVTEQEEKLGMDDL 1234
Db       1259 RVRKEVDYKDDED----DDI 1274


RESULT 3
BPEA_HUMAN
ID   BPEA_HUMAN     STANDARD;      PRT;  5171 AA.
AC   O94833; Q8N1T8; Q8N8J3; Q8WXK9; Q96AK9; Q96DQ5; Q9H555;
DT   16-OCT-2001 (Rel. 40, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Bullous pemphigoid antigen 1, isoforms 6/9/10 (Trabeculin-beta)
```

```
DE   (Bullous pemphigoid antigen) (BPA) (Hemidesmosomal plaque protein)
DE   (Dystonia musculorum protein).
GN   BPAG1 OR DMH OR DT OR KIAA0728.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 6), AND TISSUE SPECIFICITY.
RC   TISSUE=Keratinocytes;
RX   MEDLINE=21839111; PubMed=11751855;
RA   Okumura M., Yamakawa H., Ohara O., Owaribe K.;
RT   "Novel alternative splicings of BPAG1 (bullous pemphigoid antigen 1)
RT   including the domain structure closely related to MACF (microtubule
RT   actin cross-linking factor).";
RL   J. Biol. Chem. 277:6682-6687(2002).
RN   [2]
RP   SEQUENCE OF 1342-5171 FROM N.A. (ISOFORM 10).
RC   TISSUE=Duodenum;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 1702-4156 FROM N.A. (ISOFORM 6).
RC   TISSUE=Brain, Placenta, and Tongue;
RA   Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA   Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA   Katsuta N., Sato K., Tanikawa M., Yamazaki M., Murakawa K.,
RA   Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Isogai T., Nishi T., Ota T.,
RA   Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
RA   Sugiyama T., Irie R., Otsuki T., Sato H., Kikuchi H., Masuho Y.,
RA   Nagai K.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 3535-5171 FROM N.A. (ISOFORM 6).
RC   TISSUE=Brain;
RX   MEDLINE=22158633; PubMed=12168954;
RA   Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT   "Construction of expression-ready cDNA clones for KIAA genes: manual
RT   curation of 330 KIAA cDNA clones.";
RL   DNA Res. 9:99-106(2002).
RN   [5]
RP   SEQUENCE OF 4107-5171 FROM N.A. (ISOFORM 6).
RC   TISSUE=Brain;
RX   MEDLINE=99087487; PubMed=9872452;
RA   Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XI.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 5:277-286(1998).
RN   [6]
RP   SEQUENCE OF 4031-5171 FROM N.A. (ISOFORM 9).
RA   Smith M.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   DISEASE.
RC   TISSUE=Keratinocytes;
RX   MEDLINE=92011493; PubMed=1717441;
RA   Sawamura D., Li K., Chu M.-L., Uitto J.;
RT   "Human bullous pemphigoid antigen (BPAG1). Amino acid sequences
RT   deduced from cloned cDNAs predict biologically important peptide
RT   segments and protein domains.";
RL   J. Biol. Chem. 266:17784-17790(1991).
CC   -!- FUNCTION: Cytoskeletal linker protein. Anchors keratin-containing
CC       intermediate filaments to the inner plaque of hemidesmosomes. The
CC       proteins may self-aggregate to form filaments or a two-dimensional
CC       mesh (By similarity).
CC   -!- SUBUNIT: Homodimer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=10;
CC       Name=6; Synonyms=EA;
CC         IsoId=O94833-2; Sequence=Displayed;
CC       Name=1;
CC         IsoId=Q03001-1; Sequence=External;
CC       Name=2;
CC         IsoId=Q03001-2; Sequence=External;
CC       Name=3; Synonyms=1e;
CC         IsoId=Q03001-3; Sequence=External;
CC       Name=4;
CC         IsoId=Q03001-4; Sequence=External;
CC       Name=5;
CC         IsoId=Q03001-5; Sequence=External;
CC       Name=7; Synonyms=EB;
CC         IsoId=Q8WXK8-2; Sequence=External;
CC       Name=8;
CC         IsoId=Q03001-6; Sequence=External;
CC       Name=9;
CC         IsoId=O94833-3; Sequence=VSP_005068, VSP_005069;
CC       Name=10;
CC         IsoId=O94833-1; Sequence=VSP_005066, VSP_005067;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Highly expressed in skeletal muscle and
CC       cultured keratinocytes.
CC   -!- DISEASE: BPAG1 is an autoantigen of bullous pemphigoid
CC       [MIM:600088], an autoimmune subepithelial skin blistering disease.
CC   -!- SIMILARITY: Belongs to the plakin or cytolinker family.
CC   -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC   -!- SIMILARITY: Contains 1 SH3 domain.
CC   -!- SIMILARITY: Contains 29 spectrin repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF400226; AAL62061.1; -.
DR   EMBL; BC016991; AAH16991.1; -.
DR   EMBL; AK055189; BAB70870.1; ALT_INIT.
DR   EMBL; AK094883; BAC04449.1; ALT_INIT.
DR   EMBL; AK096713; BAC04848.1; ALT_INIT.
DR   EMBL; AB018271; BAA34448.2; -.
DR   EMBL; AL137008; CAC12899.1; -.
DR   HSSP; P02631; 1RRO.
DR   Genew; HGNC:1090; BPAG1.
DR   MIM; 113810; -.
DR   MIM; 600088; -.
DR   GO; GO:0005200; F:structural constituent of cytoskeleton; ISS.
DR   GO; GO:0045104; P:intermediate filament cytoskeleton organiza. . .; ISS.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR003108; GAS2.
DR   InterPro; IPR002017; Spectrin.
```

```
DR   Pfam; PF00036; efhand; 2.
DR   Pfam; PF02187; GAS2; 1.
DR   Pfam; PF00018; SH3; 1.
DR   Pfam; PF00435; spectrin; 28.
DR   ProDom; PD000012; EF-hand; 1.
DR   SMART; SM00054; EFh; 2.
DR   SMART; SM00243; GAS2; 1.
DR   PROSITE; PS00018; EF_HAND; 2.
KW   Antigen; Actin-binding; Coiled coil; Repeat; SH3 domain;
KW   Structural protein; Cytoskeleton; Cell adhesion; Calcium;
KW   Calcium-binding; Alternative splicing.
FT   REPEAT      264    341       SPECTRIN 1.
FT   REPEAT      349    444       SPECTRIN 2.
FT   REPEAT      450    550       SPECTRIN 3.
FT   DOMAIN      564    616       SH3.
FT   REPEAT      935   1031       SPECTRIN 4.
FT   REPEAT     1481   1581       SPECTRIN 5.
FT   REPEAT     1715   1814       SPECTRIN 6.
FT   REPEAT     1930   2008       SPECTRIN 7.
FT   REPEAT     2071   2174       SPECTRIN 8.
FT   REPEAT     2184   2282       SPECTRIN 9.
FT   REPEAT     2294   2382       SPECTRIN 10.
FT   REPEAT     2404   2502       SPECTRIN 11.
FT   REPEAT     2513   2611       SPECTRIN 12.
FT   REPEAT     2839   2940       SPECTRIN 13.
FT   REPEAT     2950   3047       SPECTRIN 14.
FT   REPEAT     3059   3156       SPECTRIN 15.
FT   REPEAT     3168   3268       SPECTRIN 16.
FT   REPEAT     3277   3376       SPECTRIN 17.
FT   REPEAT     3386   3484       SPECTRIN 18.
FT   REPEAT     3495   3594       SPECTRIN 19.
FT   REPEAT     3716   3816       SPECTRIN 20.
FT   REPEAT     3825   3925       SPECTRIN 21.
FT   REPEAT     3935   4035       SPECTRIN 22.
FT   REPEAT     4044   4142       SPECTRIN 23.
FT   REPEAT     4153   4250       SPECTRIN 24.
FT   REPEAT     4262   4362       SPECTRIN 25.
FT   REPEAT     4372   4470       SPECTRIN 26.
FT   REPEAT     4480   4578       SPECTRIN 27.
FT   REPEAT     4596   4684       SPECTRIN 28.
FT   REPEAT     4732   4790       SPECTRIN 29.
FT   CA_BIND    4798   4810       EF-HAND 1 (POTENTIAL).
FT   CA_BIND    4834   4846       EF-HAND 2 (POTENTIAL).
FT   DOMAIN      126    160       COILED COIL (POTENTIAL).
FT   DOMAIN      312    377       COILED COIL (POTENTIAL).
FT   DOMAIN      402    447       COILED COIL (POTENTIAL).
FT   DOMAIN      511    546       COILED COIL (POTENTIAL).
FT   DOMAIN      706    813       COILED COIL (POTENTIAL).

  Query Match             3.3%;  Score 239.5;  DB 1;  Length 5171;
  Best Local Similarity   18.5%;  Pred. No. 0.006;
  Matches  263;  Conservative  241;  Mismatches  448;  Indels  467;  Gaps   66;

Qy         22 ISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLALHLASDFFLKHPGKDVRLLVA 81
              : ::  |  |: |:|: ::|:| |::| :|       |     ||  |      |
Db       2395 VKEEYLKAELSRQLEGILKSFKDVEQKAE-------NHVQHLQSACASSHQ--------- 2438

Qy         82 CCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSY 141
                                      |  ::|  :    |||  : |            ||:
Db       2439 -----------------------FQQMSRDFQAWLDTKKEEQN-----------KSH 2461

Qy        142 NICFELEDSNEIFTQLYRTLFSVINNGHN------QKVHMHMVDLMSSIICEGDTV---- 191
                |  :|:              | |:|  :  :            : ||:       | ||: :
Db       2462 PISAKLD-----------VLESLIKDHKDFSKTLTAQSHMY-----EKTIAEGENLLLKT 2505

Qy        192 -------SQELLDTVLVNLVPAHKNL----NKQAYDLAKAL-LKRTAQAIEPYITTFFNQ 239
                     | |:|:  |       :|  :      ||      | |||  |      : :|:|     |
Db       2506 QGSEKAALQLQLNTIKTNWDTFNKQVKERENKLKESLEKALKYKEQVETLWPWIDKCQNN 2565

Qy        240 VLML--------GKTSIS-------DLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSN 284
              :  :          |: ||:         :: :|  |  :: || ||  |   :: ||||   ::: : ::
Db       2566 LEEIKFCLDPAEGENSIAKLKSLQKEMDQH-FGMV-ELLNNTANSLLSVC-EIDKGVVTD 2622

Qy        285 DNEERLQVVKLLAKMFGAKD------------SELASQNKPLWQCYLGRFNDIHVPI-- 329
              :|:  :| |  :: :   :|                |:: ::|   ||    :  |||  :
Db       2623 ENKSLIQKVDMVTEQLHSKKFCLENMTQKFKEFQEVSKESKRQLQCAKEQL-DIHDSLGS 2681

Qy        330 ---------RLECVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKK 380
                       |:  :  :    |  :   ||||  |  :  |  |  :  |  :    ||::  :  | |::
Db       2682 QAYSNKYLTMLQTQQKSLQALKHQVDLAKRLAQDLVVEASDSKGT--SDVLLQVETIAQE 2739

Qy        381 DILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKL 440
                                                       |  :   :: :::: ||
Db       2740 ----------------------------------------HSTLSQQVDEKCSFLETKL 2758

Qy        441 LHI-YYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYLYATLD--LNAVKALNEMWK 497
                | :::|:|          :|:|                    :|  |  |:::  :
Db       2759 QGIGHFRNTI------REMFSQ------------------FAEFDDELDSMAPVG---- 2789

Qy        498 CQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLEDDE 557
                         :|   | ||  :|    :||    |:   :   :   :   |     |      :    |
Db       2790 ---------RDAETLQKQKET---IKAFLKKLEALMASNDNANKTCKMMLATEETSPDLV 2837

Qy        558 KIRKQLEVLVSPTCSC---------KQAEGCVREITKKLGNPKQPTNPFLEMIKFLLERI 608
                |::  ||  |  |   |:         :|  ||  :: :   |:                ||:|
Db       2838 GIKRDLEAL-SKQCNKLLDRAQAREEQVEGTIKRLEEFYSKLKE--------FSILLQRA 2888

Qy        609 -------APVHIDTESISALI-------KQVNKSIDGTADDED---EGV-------PTDQ 644
                     ||  ::||:|:  :            |:   :  :  |   |  :    :|:           :  |
Db       2889 EEHEESQGPVGMETETINQQLNMFKVFQKEEIEPLQGKQQDVNWLGQGLIQSAAKSTSTQ 2948

Qy        645 AIRAGLELLKV--LSFTHPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEED 702
                :    |: :    :    :: :|:  |:||  |  :   |    |: |    :|    :
Db       2949 GLEHDLDDVNARWKTLNKKVAQRAAQLQEALLHCGRFQD--ALESLLSWMVDTEELVANQ 3006

Qy        703 FPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKET------QFAQIFEPLHK--- 753
                |       |     |   |   :: |        :       |: |           :  |    ||  |
Db       3007 KP-------PSAEFKVVKAQIQEQKLLQRLLDDRKSTVEVIKREGEKIATTAEPADKVKI 3059

Qy        754 ----SLDPSNLEHLITPLVT----IGHIALLAPDQF---AAPWKSWVATFIVKDLLMNDR 802
                    || |   |  |:    |       :   |:::|  ||       |     |:  :   |:|
Db       3060 LKQLSLLDSRWEALLNKAETRNRQLEGISVVA-QQFHETLEPLNEWLTT--IEKRLVNCE 3116

Qy        803 LPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSK------SGTSTLRLLTT 856
                | : :||    :|:::      ::     |                    ||:|     |:|::|
Db       3117 PIGTQASKL---EEQIAQHKALEDDII-----------NHNKHLHQAVSIGQSLKVLSS 3161

Qy        857 ILHSDGDLTEQGKISKPDMSR-----LRLAAGSAIVKLAQEPCYHEIITLEQYQLC--AL 909
                  | |: :    || |  |:          ::  :  |    |  |   |   :|     :: :|
Db       3162 --REDKDMVQ----SKLDFSQVWYIEIQEKSHSRSELLQQALCNAKIFGEDEVELMNWLN 3215

Qy        910 AINDEC--YQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNIN 967
                ::|:       |:    :  |  |  |  ||:              :: :   |: :   | |:   :
Db       3216 EVHDKLSKLSVQDYSTEGLWKQQSELRV-----------LQEDILLRKQNVDQALLNGL- 3263

Qy        968 VRREYLKQHAA-----VSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLK-DVKECL 1021
                  |  |||            :  :||  ::    |                        :||  :|   ||   :  |
Db       3264 ---ELLKQTTGDEVLIIQDKLEAIKARY------------------KDITKLSTDVAKTL 3302

Qy       1022 WFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKST 1081
                |:   :|:   :::|                                     |:  |   |         :   |
Db       3303 EQALQ--LARRLHSTH------------------------EELCTWLDKVEVELLSYET 3335

Qy       1082 TYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSA 1141
                           ||      :|                    | |:|
Db       3336 ---------QVLKGEEASQAQMR--------PKELK---------------------- 3354

Qy       1142 GKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKR 1201
                |:::    : ::::: ||:         |  |  :    |:   :|     :|| |   :  |       :| ::
Db       3355 -KEAKNNKALLDSLNEVSSALLELVPWRAREGLE--KMVAEDNERYRLVSDTITQKVEEV 3411

Qy       1202 DDSDLVRSELEKPRGRKKTPVTEQEEKL-GMDDLTKLVQEQKPKGSQ----------RSR 1250
                | : |      : ::      : : :||  |:||   |:  : :|  |:|            |     | |
Db       3412 DAAILRSQQFDQAADAELSWITETEKKLMSLGDI-RLEQDQTSAQLQVQKTFTMEILRHR 3470

Qy       1251 -------KRGH---TASESDEQQWPEEKRLKEDILENED 1279
```

```
                    | ||    ||    :|:|     |:  : :|:| |
Db       3471 DIIDDLVKSGHKIMTACSEEEKQ--SMKKKLDKVLKNYD 3507

RESULT 4
MLP1_YEAST
ID   MLP1_YEAST     STANDARD;      PRT;  1875 AA.
AC   Q02455;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin-like protein MLP1.
GN   MLP1 OR YKR095W OR YKR415.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=93247549; PubMed=8483450;
RA   Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT   "A new yeast gene with a myosin-like heptad repeat structure.";
RL   Mol. Gen. Genet. 237:359-369(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94205265; PubMed=8154186;
RA   Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA   Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT   "The complete sequence of a 15,820 bp segment of Saccharomyces
RT   cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
RT   new open reading frames.";
RL   Yeast 9:1349-1354(1993).
CC   -!- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC       REPAIR.
CC   -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC   -!- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L01992; AAA34783.1; -.
DR   EMBL; X73541; CAA51948.1; -.
DR   EMBL; Z28320; CAA82174.1; -.
DR   PIR; S38173; S38173.
DR   SGD; S0001803; MLP1.
DR   GO; GO:0005635; C:nuclear membrane; IDA.
DR   GO; GO:0005654; C:nucleoplasm; IDA.
DR   GO; GO:0006606; P:protein-nucleus import; IDA.
KW   Coiled coil; DNA repair.
FT   DOMAIN       69    487       COILED COIL (POTENTIAL).
FT   DOMAIN      531   1678       COILED COIL (POTENTIAL).
FT   DOMAIN     1834   1866       COILED COIL (POTENTIAL).
FT   CONFLICT    301    301       R -> A (IN REF. 1).
SQ   SEQUENCE   1875 AA;  218455 MW;  683A0D34C9066867 CRC64;

  Query Match             3.3%;  Score 235.5;  DB 1;  Length 1875;
  Best Local Similarity   19.5%;  Pred. No. 0.0024;
  Matches  301;  Conservative  246;  Mismatches  602;  Indels  391;  Gaps   72;

Qy         19 VKEISDKISKE------EMVRRLKMVVKTFMDMDQDSEEEKEL-YLNLALHLASDFFLKH 71
              :| :|| :: |        ||  : ::|      ::   ||    :  || |  :| |   :
Db        278 IKGLSDSLNSEKQEFSAEMSLKQRLVDLLESQLNAVKEELNSIRELNTAKVIADDSKKQT 337

Qy         72 PG-----KDVRLL---VACCLADIFRI-----YAPEAPYTSPDKLKDIFMFITRQLKGLE 118
              |      |:::|     :| |  :  |:          | |         |   |:|: :||
Db        338 PENEDLLKELQLTKEKLAQCEKECLRLSSITDEADEDNENLSAKSSSDFIFLKKQLIKER 397

Qy        119 DTK---SPQFNRYFYLLEN-IAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVH 174
               ||     |   :   .||: :  : |:    |  |  |         |      :|  | ||
Db        398 RTKEHLQNQIETFIVELEHKVPIINSFK---ERTDMLENELNNAALLLEHTSNEKNAKVK 454

Qy        175 MHMVDLMSSIICEGD--TVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPY 232
                      : || |  |::::  ||  |     :        |  :  |        |  |  |
Db        455 ELNAKNQKLVECENDLQTLTKQRLD--LCRQIQYLLITNSVSNDSKGPLRKEEIQ----- 507

Qy        233 ITTFFNQVLMLGKTSI--SDLSEHVFDLILELYNI-----DSHLLLSVLPQLEFKLKSND 285
                |  ::     ::|  ||  :  |  :  ::|  ||        :   || |:  |  ||:| :
Db        508 ---FIQNIMQEDDSTITESDSQKVVTERLVEFKNIIQLQEKNAELLKVVRNLADKLESKE 564

Qy        286 NEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHP 345
               :  :  : |:          :||   :: |                :   : ::  |  :    |
Db        565 KKSKQSLQKI-------ESETVNEAK----------EAIITLKSEKMDLESRI----E 601

Qy        346 DLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDI-LLVND---HLLNFVERERTLDKR 401
              :| |:| | ||  |     |:|    :| :   :|  |:|:     |   |     :   || |  :
Db        602 ELQKELEE-LKT-SVPNEDASYSNVTIKQLTETKRDLESQVQDLQTRISQITREST-ENM 658

Qy        402 WRVRKEAMMGLAQIY-KKYALQSAAGKDAAKQI-AWIKDKLLHIYYQNSIDDRLLVERIF 459
                :  ||      :  :|  |    |   :          ||:  : :|  |   :  |||     |::|
Db        659 SLLNKE----IQDLYDSKSDISIKLGKEKSSRILAEERFKLL----SNTLD--------- 701

Qy        460 AQYMVPHNLETTERMKCLYYLYATL---DLNAVKALNEMWKCQNLLRHQVKDLLDLIKQP 516
                          |  :   |     ||  |:     |        :  |||     |:: |      :||:|
Db        702 -----LTKAENDQLRKRFDYLQNTILKQDSKTHETLNEYVSCKSKLSIVETELLNL---- 752

Qy        517 KTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCK-- 574
              | :   ::         |  :       |     | |   | ::           |:  :| |: |          :| |
Db        753 KEEQKLRVHLEKNLKQELNKLSPEK--DSLRIMVTQLQTLQKEREDLLEETRKSCQKKID 810

Qy        575 QAEGCVREITKKLGNP-------KQPTNPFLEMIKFLLERIAPVH----------IDTE 616
              : |   : |: |:                  ::    |   :|   :   :|   :             | |
Db        811 ELEDALSELKKETSQKDHHIKQLEEDNNSNIEWYQNKIEALKKDYESVITSVDSKQTDIE 870

Qy        617 SISALIKQVNKSID------GTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETF 670
                 :  :| :  | |:         | :         ||  :  : :|   ||    |:   :  |  :
Db        871 KLQYKVKSLEKEIEEDKIRLHTYNVMDETI-NDDSLRKELEKSKI-NLTD--AYSQIKEY 926

Qy        671 ESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKS----------- 718
              : |                |    |    :  |  ||::|  |       :   :   :      |:
Db        927 KDLY-----------ETTSQSLQQTNSKLDESFKDFTNQIKNLTDEKTSLEDKISLLKEQ 975

Qy        719 -----------KKGPPRQA---KYAIHC-------IHAIFSSKETQFAQI-----FEPLH 752
                            |||   ::        |    |               : |:  |   |:: ::|         : ::
Db        976 MFNLNNELDLQKKGMEKEKADFKKRISILQNNNKEVEAVKSEYESKLSKIQNDLDQQTIY 1035

Qy        753 KSLDPSNLEHLITPLVTIGHIALLAPDQFAAPWKSWVATF-IVKDLLMNDRLPGKKTTKL 811
               :      :| |    :      :               :|         :|   |   |   :  :| |  |        |:  |
Db       1036 ANTAQNNYEQELQKHADVSKTISELREQLHT-YKGQVKTLNLSRDQLENAL---KENEKS 1091

Qy        812 WVPDEEVSPE------------TMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTIL 858
                |    :|    |                    |     :     :  |:| :|            :| |:
Db       1092 WSSQKESLLEQLDLSNSRIEDLSSQNKLLYDQIQIYTAADKEVNNSTNGPGLNNILITLR 1151

Qy        859 HSDGDLTEQGKISKPDMSRLRLAAGSAIVKL--AQEPCYHEIITLEQYQLCALAINDECY 916
                     |   :  ::: |     ||          |:|   |:     :     :   |  :         `:|
Db       1152 RERDILDTKVTVAERDAKMLRQKISLMDVELQDARTKLDNSRVEKENHSSIIQQHDD--- 1208

Qy        917 QVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREY---- 972
                           : : |::| |                     ::|      |    |   |   |  | ::|
Db       1209 ---------IMEKLNQLNL--------------LRESNITLRNELENNNNKKKELQSEL 1244

Qy        973 --LKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEI--- 1027
                 |||: |   |   |: |                       |    :   ::||   ||          |:
Db       1245 DKLKQNVAPIESELTAL----------------KYSMQEKEQELKLAKE------EVHRW 1282

Qy       1028 ------LMAKNENNSHAFIRKM---VENIK---QTKDAQGPD-DAKMN-------EKLYT 1067
                     :: |:|   | :    |:    :||:|   | :  ||  :   : |  | |  |            |:| |
Db       1283 KKRSQDILEKHEQLSSSDYEKLESEIENLKEELENKERQGAEAEEKFNRLRRQAQERLKT 1342

Qy       1068 VCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKN-FSNTKNYLPPEMKSFFTPGKPK 1126
```

```
                    : :   | |  : | :|              ||   |: :      ::         :
Db       1343 ------SKLSQDSLTEQVNSLRDA-----------KNVLENSLSEANARIEELQNAKVAQ 1385

Qy       1127 TTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENED 1186
                | | |: |     | | |:   :::|       |   |    | |          ||:
Db       1386 GNNQLEAIRKLQEDAEKASRELQAKLE-------ESTTSYESTINGL---------NEE 1428

Qy       1187 YTMSSPLPGKKSDKRDDSDLVRSELEKPRG-RKKTPVTEQEEKLGMDDLTKLVQEQKPKG 1245
                |                      :: |:|| |  :::   |     |:   :||: :|: |
Db       1429 IT-----------------TLKEEIEKQRQIQQQLQATSANEQ---NDLSNIVESMK--- 1465

Qy       1246 SQRSRKRGHTASESDEQQWPEEK--RLKEDILENEDEQNSPP-------KKGKRGRPPKP 1296
                           : | |: :: :||   : | ||| ::  | |        ||        :
Db       1466 ---------KSFEEDKIKFIKEKTQEVNEKILEAQERLNQPSNINMEEIKKKWESEHEQE 1516

Qy       1297 LGGGTPKEEPTMK------TSKKGSK----KKSGPPAPEEEEEEER-----QSG------ 1335
               :     : |  :|       | :| :|     ||       ||: |||     |||
Db       1517 VSQKIREAEEALKKRIRLPTEEKINKIIERKKEELEKEFEEKVEERIKSMEQSGEIDVVL 1576

Qy       1336 ------NTEQKSKSKQHRVSRRAQQRAES-PESSAIESTQ 1368
                    ::| |  ::  ::: |:  :  | || |   :
Db       1577 RKQLEAKVQEKQKELENEYNKKLQEELKDVPHSSHISDDE 1616
```

```
RESULT 5
CENE_HUMAN
ID   CENE_HUMAN     STANDARD;      PRT;  2663 AA.
AC   Q02224;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Centromeric protein E (CENP-E protein).
GN   CENPE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93024922; PubMed=1406971;
RA   Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT   "CENP-E is a putative kinetochore motor that accumulates just before
RT   mitosis.";
RL   Nature 359:536-539(1992).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=95196755; PubMed=7889940;
RA   Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT   "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT   microtubule motor.";
RL   EMBO J. 14:918-926(1995).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=98437347; PubMed=9763420;
RA   Chan G.K.T., Schaar B.T., Yen T.J.;
RT   "Characterization of the kinetochore binding domain of CENP-E reveals
RT   interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL   J. Cell Biol. 143:49-63(1998).
CC   -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC       KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC       OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC       AND/OR SPINDLE ELONGATION.
CC   -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC   -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CC       CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC       QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z15005; CAA78727.1; -.
DR   PIR; S28261; S28261.
DR   HSSP; P17119; 3KAR.
DR   Genew; HGNC:1856; CENPE.
DR   GK; Q02224; -.
DR   MIM; 117143; -.
DR   GO; GO:0005699; C:kinetochore; TAS.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0008350; F:kinetochore motor activity; TAS.
DR   GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
DR   GO; GO:0007079; P:mitotic chromosome movement; TAS.
DR   GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW   Cell cycle; Centromere.
FT   DOMAIN        1    335       KINESIN-MOTOR.
FT   DOMAIN      336   2471       COILED COIL (POTENTIAL).
FT   DOMAIN     2472   2663       GLOBULAR (POTENTIAL).
FT   NP_BIND      86     93       ATP (BY SIMILARITY).
SQ   SEQUENCE   2663 AA;  312087 MW;  CEFC13880C8C8CB8 CRC64;

  Query Match             3.2%;  Score 229;  DB 1;  Length 2663;
  Best Local Similarity   18.2%;  Pred. No. 0.0077;
  Matches  275;  Conservative  268;  Mismatches  537;  Indels  434;  Gaps   64;

Qy         17 PGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEE----------EKELYLNLALHLASD 66
              | | |:|    : | :::| :    |  ||: :  ||               ||:    ||  |
Db        333 PYVNEVS---TDEALLKRYR---KEIMDLKKQLEEVSLETRAQAMEKD---QLAQLLEEK 383

Qy         67 FFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIF-MFITRQLKGL-EDTKSPQ 124
                |:                            :|::::  | :|       | :: |: :
Db        384 DLLQK-------------------------VQNEKIENLTRMLVTSSSLTLQQELKAKR 417

Qy        125 FNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSSI 184
                  |  :  |   |  :|: |                |   |::   |     | |   ::|:  |
Db        418 KRRVTWCLGKINKMKNSN---------------YADQFNIPTN-ITTKTHKLSINLLREI 461

Qy        185 ---ICEGDTVSQELLDTVL-VNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTFFNQV 240
                 :|    |      |||:   :   || | ||::   :
Db        462 DESVCSESDVFSNTLDTLSEIEWNPATKLLNQENIE----------------------- 497

Qy        241 LMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMF 300
                    : :: |         :|:|        |     |    :  :| |||  :: :   :  :  :   |
Db        498 -----SELNSLRADYDNLVL-----DYEQLRTEKEEMELKLKEKNDLDEFEALERKTK-- 545

Qy        301 GAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSH 360
                 || |:                                      |::      |:|  :: :| :
Db        546 --KDQEMQ------------------------------LIHEISNLKNLVKHREVYNQ 571

Qy        361 DPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVREERTLDKRWRVRKEAMMGLAQIYKKYA 420
              | |   :              ::| ::|    :   :  ::|     ::      :    | |:        |:
Db        572 DLENEL----------SSKVELLREKEDQIKKLQEYIDSQK---LENIKMDLS-----YS 613

Qy        421 LQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYL 480
              |:|         ||  :  :  :                      |  :  :  ::    |||  |:|| |
Db        614 LESIEDPKQMKQTLFDAETV------------ALDAKRESAFLRSENLELKEKMKELATT 661

Qy        481 YATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITR------ 534
              |   ::         |::     |:  |   :|     :|| |:              ::: |::: :|
Db        662 YKQME-------NDIQLYQSQLEAKKKMQVDLEKE------LQSAFNEITKLTSLIDGKV 708

Qy        535 ------NLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLG 588
                        ||    ||   |  |:   :  :|::|  :|:::  :|:|        |     |  : |::
Db        709 PKDLLCNLELEGKITDLQKELNKEVEENEALREEV-ILLS---ELKSLPSEVERLRKEI- 763
```

```
Qy        589 NPKQPTNPFLEMIKFLLERIAPVHIDTES-ISALIKQVNKSIDGTADDEDEGVPTDQAI- 646
              |  :  | :|     :::    :  ||  :  |:::: |: |  |   :     |||
Db        764 ---QDKSEELHIITSEKDKLFSEVVHKESRVQGLLEEIGKTKDDLATTQSNYKSTDQEFQ 820

Qy        647 --------------------RAGLELLKVLSFTHPISFHSAETFESLLACLKMD----- 680
                                  |   |:: :                |: |:| |  || :
Db        821 NFKTLHMDFEQKYKMVLEENERMNQEIVNLSK--------EAQKFDSSLGALKTELSYKT 872

Qy        681 ---DEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKS-----------------K 719
                 || |   ::  |   :::|   :  | |  |   |:                   :
Db        873 QELQEKTREVQERL--NEMEQLKEQLENRDSPLQTVEREKTLITEKLQQTLEEVKTLTQE 930

Qy        720 KGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSN-----LEHLITPLVTIGHIA 774
              |    :| : ::             ::    :  |  : :: ::|          || |          || :
Db        931 KDDLKQLQESLQI------ERDQLKSDIHDTVNMNIDTQEQLRNALESLKQHQETINTL- 983

Qy        775 LLAPDQFAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVR 834
                              || ::  :  ::| |                     || : ||   :  |             :
Db        984 ----------KSKISEEVSRNLHM---------------EENTGETKDEFQ------Q 1010

Qy        835 WLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISK--PDMSRLRLAAGSAIVKLAQ- 891
                ::|:         :| :  ||   :   :  : :| ||     :   |:    | | :  |
Db       1011 KMVGIDKKQDLEAKNT-QTLTADVKDNEIIEQQRKIFSLIQEKNELQQMLESVIAEKEQL 1069

Qy        892 -----EPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALC 946
                     |    |  |: :|     : ||  : ::: ||: :  :  :     |    |
Db       1070 KTDLKENIEMTIENQEELRL----LGDELKKQQEIVAQEKNHAIKK---EGELSRTCDRL 1122

Qy        947 A--KDPVKERRAHARQCLVKNINVRREYLKQHAAVSE--------KLLSLLPEYVVPYTI 996
              |  :: :||:     ::    : :||: |  :    ::|             |    |  |::    :
Db       1123 AEVEEKLKEKSQQLQEKQQQLLNVQEEMSEMQKKINEIENLKNELKNKELTLEHMETERL 1182

Qy        997 HLLAH-DPDYVKVQDI-EQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQ 1054
              |     : :| :|: | :: |: :||     |:         ::     :||::         |||:
Db       1183 ELAQKLNENYEEVKSITKERKVLKE-----LQKSFETERDHLRGYIREIEATGLQTKEEL 1237

Qy       1055 GPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPP 1114
                : |   |: ::  ::  |: |                   |:        |       |:||   |
Db       1238 KIAHIHLKEHQETIDELRRSV--SEKT------------AQIINTQDLEKSHTK--LQE 1280

Qy       1115 EMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKS--SRMETVSNASSSSNPSSPGRIKG 1172
              |:          ||      :  |  :   | |:|:   : :|  ::   |:: :  ::  ||
Db       1281 EIP------------VLHEEQELLPNVKKVSETQETMNELELLTEQSTTKDSTTLARI-- 1326

Qy       1173 RLDSSEMDHSE-NEDYTMSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTE------- 1224
                      ||:       || :   |          : :||:    ::    ||      :  |   :  |
Db       1327 -----EMERLRLNEKFQESQEEIKSLTKERDNLKTIKEALEVKHDQLKEHIRETLAKIQE 1381

Qy       1225 ----QEEKLGM----DDLTKLVQEQ---KPK-----------GSQRSRKRGH-----TA 1256
                  ||: |  |        :: ||:| |        |||                |   :   :       |
Db       1382 SQSKQEQSLNMKEKDNETTKIVSEMEQFKPKDSALLRIEIEMLGLSKRLQESHDEMKSVA 1441

Qy       1257 SESDEQQ------WPEEKRLKEDI-------LENEDEQNSPPKKGKRGRPP-KPLGGGTP 1302
                 |  |: |       |   :|||:|         || |:|                  |       |
Db       1442 KEKDDLQRLQEVLQSESDQLKENIKEIVAKHLETEEELKVAHCCLKEQEETINELRVNLS 1501

Qy       1303 KEEPTMKTSKK-----GSKKKSGPPAPEEEEEE---ERQSGNTEQKSKSKQHRVSRRAQQ 1354
              ::|   :  |  :|          |  ::       |:||:    :: |     |   :: ||  :   |:|:
Db       1502 EKETEISTIQKQLEAINDKLQNKIQEIYEKEEQLNIKQISEVQENVNELKQFKEHRKAKD 1561

Qy       1355 RA-ESPESSAIEST 1367
                 | :| || :| |
Db       1562 SALQSIESKMLELT 1575
```

```
RESULT 6
RIF1_YEAST
ID   RIF1_YEAST     STANDARD;      PRT;  1916 AA.
AC   P29539;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   RIF1 protein (RAP1-interacting factor 1).
GN   RIF1 OR YBR275C OR YBR1743.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB320;
RX   MEDLINE=92249772; PubMed=1577274;
RA   Hardy C.F.J., Sussel L., Shore D.;
RT   "A RAP1-interacting protein involved in transcriptional silencing and
RT   telomere length regulation.";
RL   Genes Dev. 6:801-814(1992).
RN   [2]
RP   SEQUENCE OF 1-1096 FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=94378722; PubMed=8091861;
RA   Holmstroem K., Brandt T., Kallesoe T.;
RT   "The sequence of a 32,420 bp segment located on the right arm of
RT   chromosome II from Saccharomyces cerevisiae.";
RL   Yeast 10:S47-S62(1994).
RN   [3]
RP   SEQUENCE OF 1005-1916 FROM N.A.
RC   STRAIN=S288c;
RA   Aigle M., Baclet M.C., Barthe C., Biteau N., Crouzet M., Doignon F.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   FUNCTION, AND SUBUNIT.
RX   MEDLINE=97242430; PubMed=9087429;
RA   Wotton D., Shore D.;
RT   "A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to
RT   regulate telomere length in Saccharomyces cerevisiae.";
RL   Genes Dev. 11:748-760(1997).
CC   -!- FUNCTION: Involved in transcriptional silencing and telomere
CC       length regulation. Its role in telomere length regulation results
CC       from either a block in elongation or promoting degradation of the
CC       telomere ends. Loss of RIF1 function results in derepression of an
CC       HMR silencer, whose ARS consensus element has been deleted, and in
CC       the elongation of telomeres. RAP1 may target the binding of RIF1
CC       to silencers and telomeres.
CC   -!- SUBUNIT: Interacts with RIF2 and RAP1 C-terminus.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z36144; CAA85238.1; -.
DR   EMBL; X76053; CAA53638.1; -.
DR   EMBL; X66501; CAA47121.1; -.
DR   PIR; S46157; S46157.
DR   SGD; S0000479; RIF1.
DR   GO; GO:0005696; C:telomere; IDA.
DR   GO; GO:0006348; P:chromatin silencing at telomere; IMP.
DR   GO; GO:0007004; P:telomerase-dependent telomere maintenance; IMP.
KW   Telomere.
FT   CONFLICT    580    580       C -> S (IN REF. 1).
FT   CONFLICT    732    732       T -> A (IN REF. 1).
SQ   SEQUENCE   1916 AA;  217959 MW;  17B5487151CFE665 CRC64;

  Query Match             3.2%;  Score 227;  DB 1;  Length 1916;
  Best Local Similarity   17.4%;  Pred. No. 0.0062;
  Matches  316;  Conservative  267;  Mismatches  612;  Indels  624;  Gaps   80;

Qy          3 HSKTRTNDGKITY-----------PPGVKEISDKI----------SKEEMVRRLKMV--- 38
                : :|| |   : |                    | |   :  |            |:: : || ::
Db        183 YKETRVNPRALEYWVSGEIHGLVDNESVSEFKEIIEGGLGILRQESEDYVARRFEVYATF 242

Qy         39 -----VKTFMDMDQDSEEEKELYLNLA---------LHLASDFFLKHPGKDVRLLVACCL 84
                     : |  ::::  ::    | :|:              | :| |  |     |     ::
```

```
Db        243 NNIIPILTTKNVNEVDQKFNILIVNIESIIEICIPHLQIAQDTLLSSSEKKNPFVI---- 298

Qy         85 ADIFRIYAPEAPYTS--------------PD---KLKDIF-------------MFITR 112
                  |:|     : |              ||   ||| |:              : ||
Db        299 ----RLYVQIVRFFSAIMSNFKIVKWLTKRPDLVNKLKVIYRWTTGALRNENSNKIIITA 354

Qy        113 QLKGLEDTKSPQFNRYFYLLENI-AWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQ 171
              |:   | | |    |   :|    | |    : ::     |:    | |: :|     :: |
Db        355 QVSFLRDEK---FGTFFLSNEEIKPIISTFTEIMEINSHNLIYEKL------LLIRGFLS 405

Qy        172 KVHMHMVDLMSS---------IICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALL 222
              |    |:: ::|          ||    :   | ::| | :| |:             ||
Db        406 KYPKLMIETVTSWLPGEVLPRIIIGDEIYSMKILITSIVVLL--------------ELL 450

Qy        223 KRTAQAIEPYITTFFNQVLMLGKTSISDLSEHVFDLILELY---NID----SHLLLSVLP 275
              |:     ::    | :||         :     :   |    | |   |:|     |||   :
Db        451 KKCLDFVDEHERIY--QCIMLSPVCETIPEKFLSKLPLNSYDSANLDKVTIGHLLTQQIK 508

Qy        276 QLEFKLKSNDNEERLQV-VKLLAKMF--GAKDSELASQNKPLW-----QCYLGRFNDIHV 327
                : :   |||:  : : : :   ::   |  :  :| |:: :|      |::      |
Db        509 --NYIVVKNDNKIAMDLWLSMTGLLYDSGKRVYDLTSESNKVWFDLNNLCFINN----HP 562

Qy        328 PIRLECVK---FASHCL-----MNHPDLAKDLTEYLK---------VRSHDPEEAIRHDV 370
                || :|      ::|:        :  :   |  |    |:           |        |  |  :  :
Db        563 KTRLMSIKVWRIITYCICTKISQKNQEGNKSLLSLLRTPFQMTLPYVNDPSAREGIIYHL 622

Qy        371 IVSIVTA--AKKDIL-----LVNDHLLNFVRERTLDKRWRVR---------------KEA 408
              :  : ||  : |::        |  |||:  :  |  :  |     :                  | |
Db        623 LGVVYTAFTSNKNLSTDMFELFWDHLITPIYEDYVFKYDSIHLQNVLFTVLHLLIGGKNA 682

Qy        409 MMGLAQIYKKY-------ALQSAAGKDAAK---QIAWIK---DKLLHIYYQNSIDDRLLV 455
               : |  :  |||:           |  :   ||   :     ||  ||    ||:: :  :|         |
Db        683 DVALERKYKKHIHPMSVIASEGVKLKDISSLPPQI--IKREYDKIMKVVFQT-------V 733

Qy        456 ERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQ 515
              |    :    : |:|  |                                  |:|   :|
Db        734 EVAISNVNLAHDLILTS----------------------------LKHLPED------- 757

Qy        516 PKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVL-----------EDDEKIRKQLE 564
               |    :::   | : :|:|  |      :||      |                 |    :    ::
Db        758 RKDQTHLESFSSLILKVTQNNKDTPIFRDFFGAVTSSFVYTFLDLFLRKNDSSLVNFNIQ 817

Qy        565 VLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLL--------------ERIA 609
              :                 :::: :|    |:  ||  : |||                      :
Db        818 ISKVGISQGNMTLDLLKDVIRKARN---ETSEFLIIEKFLELDDKKTEVYAQNWVGSTLL 874

Qy        610 PVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAET 669
              | :|        :|   |||                   || :  :|      ||| |::
Db        875 PPNISFREFQSLANIVNK-----------VPNENSIENFLDLCLKLSF--PVNL----- 915

Qy        670 FESLLACLKMDDEKVAEAALQIFKNTGSKIEEDF----------PHIRSALLPVLHHKS 718
              |  |    :  ::   :           : ||  :|:  |                |  : | ||| |
Db        916 FTLLHVSMWSNNNFIYFIQSYVSKNE-NKLNVDLITLLKTSLPGNPELFSGLLPFL---- 970

Qy        719 KKGPPRQAKY---AIHCIHA-------------------------------IFSSKE- 741
                   |: |:     :|||:                                                         :|     |
Db        971 -----RRNKFMDILEYCIHSNPNLLNSIPDLNSDLLLKLLPRSRASYFAANIKLFKCSEQ 1025

Qy        742 ----------------TQFAQIFEPLHKSLDPSNLE--HLITPLVTIGHIALLAPDQFA 782
                                   |::|    |   :  |  |  ||   :|  |:     |:|:    |  :
Db       1026 LTLVRWLLKGQQLEQLNQNFSEIENVLQNASD-SELEKSEIIRELL---HLAMANPIE-- 1079

Qy        783 APWKSWVATFIVKDLLMN--DRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMK 840
               |   |  :   |  :|: :  :  :        |       |   | :::  ::|||  :::|   :
Db       1080 -PLFSGLLNFCIKNNMADHLDEFCGNMTSEVLF---KISPELLLKLLTYK---------- 1125

Qy        841 NNHSKSGTSTLRLLTTILHSDGD----LTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYH 896
                     :|       ::   |   :  |  |        |  |:   |  |             |::    :||
Db       1126 --EKPNGKLLAAVIEKIENGDDDYILELLEKIIIQKE----------IQILEKLKEP--- 1170

Qy        897 EIITLEQYQLCALAINDECYQ-----VRQVFAQKLHKGLSRLRLPLEYMAICALCAKDP- 950
                  |   :  |    ::  |   :  ::          :|::      |  |  |||         :   :  ::    :|
Db       1171 ----LLVFFLNPVSSNMQKHKKSTNMLRELVLLYLTKPLSRSAAKKFFSMLISILPPNPN 1226

Qy        951 --------------------VKERRAHARQCLVKNINVRREYLKQ----HAAVSEKLLSL 986
                                   |::|  :         :|  |    :::::      |      |  |  :
Db       1227 YQTIDMVNLLIDLIKSHNRKFKDKRTY--NATLKTIG---KWIQESGVVHQGDSSKEIEA 1281

Qy        987 LPE----YV---------------------VPYTIHLLAHDPDYVKVQDIEQLKDVKEC 1020
              :|:      |:                      ||  |      :   |   :::      ||   :
Db       1282 IPDTKSMYIPCEGSENKLSNLQRKVDSQDIQVPAT-QGMKEPPSSIQISSQISAKD-SDS 1339

Qy       1021 LWFVLEILMAKNENNSHA----------------FIRKMVENIKQTK-----------D 1052
              :     :|   ::   |||                         ||:: :  :|  |:
Db       1340 ISLKNTAIMNSSQQESHANRSRSIDDETLEEVDNESIREIDQQMKSTQLDKNVANHSNIC 1399

Qy       1053 AQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYL 1112
              :      |:   :  |  |:    |       :  :  :              ||       |         |  ||  |     |:  |  |
Db       1400 STKSDEVDVTE-LHESIDTQSSEVNAYQPIEVLTSELKAVTNRSIKTNPDHNVVNSDNPL 1458

Qy       1113 PPEMKSFFTPGKPKT------------------------TNVLGAVNKPLSSA------ 1141
                    |     |      ::                                ||:     |:   |  |
Db       1459 KRPSKETPTSENKRSKGHETMVDVLVSEEQAVSPSSDVICTNIKSIANEESSLALRNSIK 1518

Qy       1142 ----------------GKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSE------- 1178
                                :|:  ||          |  :             |     |    :|    :   :   :
Db       1519 VETNCNENSLNVTLDLDQQTITKEDGKGQVEHVQRQENQESMNKINSKSFTQDNIAQYKS 1578

Qy       1179 ------MDHSENEDYTMSSPLPGKKSDKRDDSDLVRSE-----LEKPRGRKKTPVTEQEE 1227
                      :   ||  ||   :                     :  :   ||:|          ::  |  |             ::|  |
Db       1579 VKKARPNNEGENNDYACNV----------EQASPVRNEVPGDGIQIPSGTILLNSSKQTE 1628

Qy       1228 KLGMDDLTK-------LVQEQKPKGSQSRSRKRGHTASESDEQQWPEEKRLKEDILEN--- 1277
               |  :|||                :  ||:       |:    ||  :  :      :|       |      ||:    :     :   |
Db       1629 KSKVDDLRSDEDEHGTVAQEKHQVGAINSRNKNNDRMDSTPIQGTEEESREVVMTEEGIN 1688

Qy       1278 ---EDEQNSPPKKGKRGRPPKPLGGG--------------TPKEEPTMKTSKKGSKKKSG 1320
                    ||            |     :|          || |                                 ::|     :|: :          |::|
Db       1689 VRLEDSGTCELNKNLKG----PLKGDKDANINDDFVPVEENVRDEGFLKSMEHAVSKETG 1744

Qy       1321 PPAPEEEEEEE------------------RQSGNTEQKSKSKQHRVSRRAQQRAESPESS 1362
                    ||: |                                 :    |:     |    :     :  |:   |              :||
Db       1745 -----LEEQPEVADISVLPEIRIPIFNSLKMQGSKSQIKEKLKKRLQRNELMPPDSPPRM 1799

Qy       1363 AIESTQSTPQKGRGRPSKT 1381
                |:|      |  |      ||
Db       1800 T-ENTNINAQNGLDTVPKT 1817
```

RESULT 7
ALM1_SCHPO
ID   ALM1_SCHPO   STANDARD;   PRT;   1727 AA.
AC   Q9UTK5; O13313; Q9UTT8;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Abnormal long morphology protein 1 (Sp8).
GN   ALM1 OR SPAC1486.04C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,

```
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
RN   [2]
RP   SEQUENCE OF 495-1727 FROM N.A., AND CHARACTERIZATION.
RC   STRAIN=972;
RX   MEDLINE=20123449; PubMed=10660053;
RA   Jimenez M., Petit T., Gancedo C., Goday C.;
RT   "The alm1+ gene from Schizosaccharomyces pombe encodes a coiled-coil
RT   protein that associates with the medial region during mitosis.";
RL   Mol. Gen. Genet. 262:921-930(2000).
RN   [3]
RP   SEQUENCE OF 644-834 FROM N.A.
RC   STRAIN=968 h90;
RX   MEDLINE=20223868; PubMed=10759889;
RA   Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
RA   Hiraoka Y.;
RT   "Large-scale screening of intracellular protein localization in living
RT   fission yeast cells by the use of a GFP-fusion genomic DNA library.";
RL   Genes Cells 5:169-190(2000).
CC   -!- FUNCTION: AFTER THE ONSET OF MITOSIS, AT MID- TO LATE ANAPHASE,
CC       CO-LOCALIZES WITH THE MEDIAL ACTIN RING. MAY PLAY A ROLE IN
CC       CYTOKINESIS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AL133357; CAB62414.1; -.
DR   EMBL; AF010473; AAB65416.1; ALT_INIT.
DR   EMBL; AB028012; BAA87316.1; -.
DR   PIR; T50073; T50073.
DR   GeneDB_SPombe; SPAC1486.04c; -.
KW   Coiled coil.
FT   DOMAIN       57    361       COILED COIL (POTENTIAL).
FT   DOMAIN      443    463       COILED COIL (POTENTIAL).
FT   DOMAIN      542    740       COILED COIL (POTENTIAL).
FT   DOMAIN      804   1106       COILED COIL (POTENTIAL).
FT   DOMAIN     1223   1427       COILED COIL (POTENTIAL).
FT   DOMAIN     1497   1555       COILED COIL (POTENTIAL).
FT   DOMAIN     1601   1664       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1727 AA;  197858 MW;  F820BF8D9C132644 CRC64;

  Query Match             3.1%;  Score 220.5;  DB 1;  Length 1727;
  Best Local Similarity   17.8%;  Pred. No. 0.011;
  Matches  266;  Conservative  251;  Mismatches  543;  Indels  433;  Gaps   61;

Qy        146 ELEDSNEIFTQLYRTLFSVINNGHNQKVHM-------------------HMVDLMSSII 185
              ||:: ::| ::          |      |:  |:                ||    || :
Db        130 ELQEQSKIASEQLSIAKDQIEALQNENSHLGEQVQSAHQALSDIEERKKQHMFASSSSRV 189

Qy        186 CEGDTVSQE-LLDTVLVNLVPAHKNL-------NKQAYDLAKAL---------LKRTAQA 228
                |   | ::  | : | :|  |  :         ::|  || | |          |    |
Db        190 KEEILVQEKSALVSDLASLQSDHSKVCEKLEVSSRQVQDLEKKLAGLAQQNTELNEKIQL 249

Qy        229 IEPYITTF-----FNQVLMLGKTSISDLSEHV--FDLILELYNIDSHLLLSVLPQLEFKL 281
               |    : :         :::|    ||| :| | |     :   |:    |  |  |  :  |: ||
Db        250 FEQKRSNYSSDGNISKILETDPTSIKELEEEVETQKRLTALWESKSSELQSEVAALQEKL 309

Qy        282 KS-----------NDNEER--------------------LQVVK------------- 294
               |              |:|:::                     |||||
Db        310 TSQQSLYNNVTEELNNNKQQLLISENSLRELQEKYDSVVSELQVVKENKNTSVSAGVGLF 369

Qy        295 --LLAKMFGAKDSELA-----SQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDL 347
                |  |:    :: |  :       |  |   |  |      : : : : :        ||:| |
Db        370 SPLAQKLSAVQNPEFSFTKVYSDNMKLQQ----KVSSLKLQLDRLTNKFSSFC------- 418

Qy        348 AKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKE 407
               : : : : |       | :|::: :: ::      : |    :::  |    |:     |     ::|:|
Db        419 -EQVKQRIPVVKQQRSEIVRNNIYMNFLS----ESLETSNNNLTKVQAELLST--KMRQE 471

Qy        408 AMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHI-----------------YYQN-S 448
              |        | :          | ::::  :   :| |:                       | || |
Db        472 A------CYLQLTASRTQCSDLSREVICLMAELDHLNETKSRNVPATVQVALDEYAQNPS 525

Qy        449 IDDRLLVERIFAQY-----MVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLR 503
                   || :   | :          |    ||  | :::  |           |:   ::         ::   |  ::
Db        526 TASETLVNKELANFSSIKEAVSKTLELREKVRAL-----ECDVE-IQKQTVQYQISNAVK 579

Qy        504 HQVKDLLDLIKQPKTDASVKAIFSKVMVITRNL---------------PDPGKAQDFMK 547
                   | : ||   ::: :     | :: ::   |||                                  |      | ||
Db        580 ENSNTLSEQIKNLESELNSSKIKNESLLNERNLLKEMLATSRSSILSHNSSAGNIDDKMK 639

Qy        548 KF---TQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVRE---ITKKLGNPKQPTNPFLEMI 601
                   |: ||  : ::  :              :   :   |  |: :     :   |    |  |:    |  |:| |
Db        640 SIDESTRELEKNYEVYRNEMTAIQESLS-KRNQDLLSEMEAIRKELENSKYQQ------- 691

Qy        602 KFLLERIAPVHIDTESISALIKQ---VNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSF 658
               :    :|:     :  |  |:         |:     :|:::     :|:          :|  :    ||| |  |
Db        692 QLSTDRLTNANNDVEAFKKEAKELRSINQNLQDIISRQDQ-----RASKFAEELLHVNSL 746

Qy        659 THPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKI---------EEDFPHIRS 708
                         ||   :    |      |  : :              :  |  |     ::                |       :  |
Db        747 --------AERLKGELNASKGEKD--------LRKRTQERLISENDKLLAERERLMSLVS 790

Qy        709 ALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKET------------QFAQIFEPLHKSL 755
               |    |: :           |:  |:               |  ||:                   :  :       |  |||
Db        791 DLQTFLNQQQLSDAARKVKFE--------SEKESLSLSLQKLKESNEKMSNDLHSLQKSL 842

Qy        756 DPSNLEHLITPLVTIGHIALLAPDQFAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPD 815
               : |  :|:                                :  :|  |::    :    :     ||     |
Db        843 EKSGIEY---------------------------SSRIKTLMLEKQSLSEDNRKLL--D 872

Qy        816 EEVSPETMVKIQAIKMMVRW------LLGMKNNHSKSGTSTLR------LLTTILHSDGD 863
               :     |      :|:|  :    ::             |         |       :    :       |     |:   :
Db        873 NQQMME--IKLQELNGVIELEKQRFSTLEAKFTQQKNTSYSEREALLESSLSDLQSKHTS 930

Qy        864 LTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFA 923
               |    |      |    :: :|: |:        |||:|              ::|:|                             :
Db        931 LESQYNYSLRNIEQLQAAS-----KLAEEMVERVKTEYDEYRLQT-------------S 971

Qy        924 QKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKL 983
               : |  |            :: |  :|            |:|         |      :  :  ::|  |  :  :           ||
Db        972 ESLEK---------NHLKITSL-------EQRIVILQDEIASSSLRCENITKD---SETR 1012

Qy        984 LSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNS-----HA 1038
              ::|| |                 |  :|       :  | :|:  | |:  |   |   |  |              |:          ||
Db       1013 VALLLEENKHLNNELSSH--RNAEKQHLEKENDYKQQLLLVTEDLRKTREDYEKELLRHA 1070

Qy       1039 FIRKMVENIKQ--TKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLP-- 1094
                |   :: :::    ||    :    :|  :  :|:              ::|:      :::|   :   :     :||:
Db       1071 DARSTLQKLREDYTKALEQVED--LNK------EIALKAGINESQPFPISEKEDPLRQEV 1122

Qy       1095 -------ARFFTQ---PDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQ 1144
                            |    ||       :  ||:    :       |::  |         |               |   :   ::|   :            |
Db       1123 YVLKKQNAMLLTQLQSSNLNFAEITS-PSPDLDSVMKLGLSDLQNHVKRISKEMEIISCQ 1181
```

```
Qy       1145 SQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSD----- 1199
                |     : :     |||         |: :        :| : :|   |::|:
Db       1182 RQLLFLENKKLKRTVESSN---------RVIADLQRGITEKDVSSTSESVGERSNYLNMV 1232

Qy       1200 --KRDDSDLVRSELEKPRGRKKTPVTEQEEK-----------LGMDDLTKLVQEQKP-- 1243
                 : :  :|  ||    |  : :|| ||           |  : |    :| :|
Db       1233 ALLNESNKSLRENLE----RNEEVITELREKIETLKTDLANFRLNKEQLESQLQTEKAAV 1288

Qy       1244 ---KGSQRSRKR----------GHTASESDEQQWPEEKRLKEDILENEDEQNSPPKKGKR 1290
                 : |     ||               |:: ||  :   |   ||:::|  :::
Db       1289 KKLENSNEEYKRHNQEILLSLNSSTSTSSDASRLKNELVSKENLIEELNQEI-------- 1340

Qy       1291 GRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEEERQSGNTE----------- 1338
                           |  |  |  ::| |  |:       |  : : |: :   ||:
Db       1341 ---------GHLKSE--LETVKSKSEDLENERAQNQSKIEQLELKNTKLAAAWRTKYEQV 1389

Qy       1339 -QKSKSKQHRVSRRAQQRAESPESSAIESTQSTPQKGRGRPSKTP---SPSQP 1387
                ||  | ::: ::  |:     |:   |  |   |     :|| ||    : |:|
Db       1390 VNKSLEKHNQIRQQLSQKTSELEAKVAECHQLNEQ--LNKPSATPTATTQSEP 1440
```

RESULT 8

```
NFH_HUMAN
ID   NFH_HUMAN      STANDARD;      PRT;  1020 AA.
AC   P12036;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neurofilament triplet H protein (200 kDa neurofilament protein)
DE   (Neurofilament heavy polypeptide) (NF-H).
GN   NEFH OR NFH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88328981; PubMed=3138108;
RA   Lees J.F., Shneidman P.S., Skuntz S.F., Carden M.J., Lazzarini R.A.;
RT   "The structure and organization of the human heavy neurofilament
RT   subunit (NF-H) and the gene encoding it.";
RL   EMBO J. 7:1947-1955(1988).
CC   -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC       AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC       NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC       SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC   -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC       PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC       THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC       INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC       OF AXONAL CALIBER.
CC   -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC       OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC       LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC       COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X15306; CAA33366.1; -.
DR   EMBL; X15307; CAA33366.1; JOINED.
DR   EMBL; X15308; CAA33366.1; JOINED.
DR   EMBL; X15309; CAA33366.1; JOINED.
DR   PIR; S00979; QFHUH.
DR   Genew; HGNC:7737; NEFH.
DR   MIM; 162230; -.
DR   GO; GO:0005883; C:neurofilament; NAS.
DR   GO; GO:0007399; P:neurogenesis; NAS.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Neurone; Phosphorylation.
FT   DOMAIN        1    100       HEAD.
FT   DOMAIN      101    413       ROD.
FT   DOMAIN      414   1020       TAIL.
FT   DOMAIN      101    132       COIL 1A.
FT   DOMAIN      133    145       LINKER 1.
FT   DOMAIN      146    244       COIL 1B.
FT   DOMAIN      245    266       LINKER 12.
FT   DOMAIN      267    288       COIL 2A.
FT   DOMAIN      289    292       LINKER 2.
FT   DOMAIN      293    413       COIL 2B.
SQ   SEQUENCE   1020 AA;  111780 MW;  1177C9DCB3DCF1D4 CRC64;

  Query Match             3.0%;  Score 217.5;  DB 1;  Length 1020;
  Best Local Similarity   19.2%;  Pred. No. 0.0074;
  Matches  206;  Conservative  162;  Mismatches  426;  Indels  281;  Gaps   44;

Qy        377 AAKKDILLVNDHLLN---FVRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGK-DAAKQ 432
              ||:  : |  :|||     ||:| ||    | |:||    |:    ::| :: | : |  |:
Db        162 AARGQLRLEQEHLLEDIAHVRQR-LDDEARQREEA-EAAARALARFAQEAEAARVDLQKK 219

Qy        433 IAWIKDK--LLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKC-----LYYLYATLD 485
                ::::   |  ::|  : : |    :                || : :||     |  : | |:
Db        220 AQALQEECGYLRRHHQEEVGELLGQIQGSGAAQAQMQAETRDALKCDVTSALREIRAQLE 279

Qy        486 LNAVKAL--NEMWKCQNLLRHQVKDLLDLIKQPKTDA------SVKAIFSKVMVITRNLP 537
               :||::   :| |    |        | |    :  |||           :        ::   |  |
Db        280 GHAVQSTLQSEEWFRVRL------DRLSEAAKVNTDAMRSAQEEITEYRRQLQARTTELE 333

Qy        538 DPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPF 597
                :| :::       |||        | |  ::      | ::|    :::: :| |  |
Db        334 ALKSTKDSLERQRSELED----RHQADI-----ASYQEA---IQQLDAELRNTKWEMAAQ 381

Qy        598 LEMIKFLLERIAPVHIDTESISALIKQVNKSIDGTADDE-----------DEGVPTDQA 645
              |    : ||      : :| |       |      | ::|    :|              ||:|   :
Db        382 LREYQDLLN--VKMALDIE-----IAAYRKLLEG---EECRIGFGPIPFSLPEGLPKIPS 431

Qy        646 IRAGL-----ELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAALQIFKN------ 694
               :    :        | :||:   :     :       |   |        :: :|:  ||  :  |
Db        432 VSTHIKVKSEEKIKVVEKSEKETVIVEEQTEETQVTEEVTEEEEKEAKEEEGKEEEGGEE 491

Qy        695 ---TGSKIEEDFPHIRSALLPVLHHKSKKGP-------PRQAKYAIHCIHAIFSSKETQF 744
                  | : |    |     |   |     |     | |         | :||            | | :
Db        492 EEAEGGEEETKSPPAEEAASP---EKEAKSPVKEEAKSPAEAK----------SPEKEE 537

Qy        745 AQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPWKSWVATFIVKDLLMNDRLP 804
              |:                                  :| :  :| |:                   : |
Db        538 AK----------------------------SPAEVKSPEKA--------------KSP 553

Qy        805 GKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDL 864
              |:   |    |  |  |||    | :|                                  |   ::
Db        554 AKEEAK--SPPEAKSPE---KEEA-----------------------------KSPAEV 578

Qy        865 TEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQ 924
                   |    |        :   |  :      : |:  |      |    :    :          : :|       :|
Db        579 KSPEKAKSPAKEEAKSPAEAKSPEKAKSPVKEEAKSPAE---AKSPVKEEAKSPAEV--- 632

Qy        925 KLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLL 984
                 |    :  : |  :   |          || | ||             |:     :  :|    |     ||
Db        633 ---KSPEKAKSPTKEEAKSPEKAKSPEKEE--------AKSPEKAKSPVKAEAKSPEKAK 681

Qy        985 SLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMV 1044
              |           :    |  |:  |      |: |   ::       |  |    ||:     :  : :::
Db        682 S--------PVKAEAKSPEKAKSPVKEEAKSPEKAKSPVKE--EAKSPEKAKSPVKEEA 730

Qy       1045 ENIKQTK-----DAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFT 1099
               : :: |           :|: |: || ||  |          :::     :: |  : |  : |
Db        731 KTPEKAKSPVKEEAKSPEKAKSPEKAKT-----LDVKSPEAKTPAKEEARSPA------- 778
```

```
Qy       1100 QPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVSNAS 1159
              ||    |: :  |:||   | | |:              || :  |  : :  : | |   |
Db        779 --DKFPEKAKSPVKEEVKS---PEKAKS---------PLKADAKAPEKEIPKKEEVK--S 822

Qy       1160 SSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKRDDSDLVRSELEKPRGRKK 1219
                  |  :| :    |        |:   ::|   :| | : :       | |  :|:
Db        823 PVKEEEKPQEVKVK----EPPKKAEEEKAPATPKTEEKKDSKKE--------EAP--KKE 868

Qy       1220 TPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASE----SDEQQWPEEKRLKEDIL 1275
                |  : |||       |     :|||  |:    |:     :      : |:: |  :  :|||
Db        869 APKPKVEEK-------KEPAVEKPKESKVEAKKEEAEDKKKVPTPEKEAPAKVEVKEDAK 921

Qy       1276 ENEDEQ--NSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEERQ 1333
                |  :      |    |    | ||                      ::||   |  :: :||: :
Db        922 PKEKTEVAKKEPDDAKAKEPSKP-----------------AEKKEAAPEKKDTKEEKAK 963

Qy       1334 SGNTEQKSKSKQHRVSRRAQQRAESPESSAIESTQSTPQKGRGRPSKTPSPSQPK 1388
                 : |:::|      :    :      |::    | :  || ||      | |       |
Db        964 KPEEKPKTEAKAKEDDKTLSKEPSKPKAEKAEKSSSTDQKDSKPPEKATEDKAAK 1018
```

RESULT 9
USO1_YEAST

```
ID   USO1_YEAST     STANDARD;      PRT;  1790 AA.
AC   P25386;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Intracellular protein transport protein USO1.
GN   USO1 OR INT1 OR YDL058W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=X2180-1A;
RX   MEDLINE=91185402; PubMed=2010462;
RA   Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA   Yamasaki M.;
RT   "A cytoskeleton-related gene, uso1, is required for intracellular
RT   protein transport in Saccharomyces cerevisiae.";
RL   J. Cell Biol. 113:245-260(1991).
RN   [2]
RP   SEQUENCE OF 782-1790 FROM N.A.
RA   Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA   Kendrick K.E.;
RL   Submitted (FEB-1993) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-8 FROM N.A.
RA   Bai Y., Symington L.S.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC       COMPLEX.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC       MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC       ER AND THE GOLGI COMPLEX.
CC   -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC       OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC       COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC   -!- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X54378; CAA38253.1; -.
DR   EMBL; L03188; AAB00143.1; -.
DR   EMBL; U53668; AAB66659.1; -.
DR   SGD; S0002216; USO1.
DR   InterPro; IPR002017; Spectrin.
DR   InterPro; IPR006955; Uso1_p115_C.
DR   InterPro; IPR006953; Uso1_p115_head.
DR   Pfam; PF04871; Uso1_p115_C; 1.
DR   Pfam; PF04869; Uso1_p115_head; 1.
KW   Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT   DOMAIN        1    724       GLOBULAR HEAD.
FT   DOMAIN      725   1790       COILED COIL (POTENTIAL).
FT   DOMAIN      465    487       CHARGED (HYPER-HYDROPHILIC).
FT   DOMAIN      991   1790       DISPENSABLE FOR THE PROTEIN FUNCTION.
FT   DOMAIN     1172   1786       ASP/GLU-RICH (ACIDIC).
FT   CONFLICT    847    847       G -> E (IN REF. 2).
FT   CONFLICT    924    924       E -> K (IN REF. 2).
FT   CONFLICT   1253   1253       V -> I (IN REF. 2).
FT   CONFLICT   1319   1319       I -> V (IN REF. 2).
FT   CONFLICT   1461   1461       N -> S (IN REF. 2).
FT   CONFLICT   1581   1581       G -> S (IN REF. 2).
FT   CONFLICT   1600   1600       I -> V (IN REF. 2).
FT   CONFLICT   1661   1661       R -> S (IN REF. 2).
FT   CONFLICT   1772   1772       D -> DEEDDEE (IN REF. 2).
SQ   SEQUENCE   1790 AA;  206424 MW;  6CE2B216E9FD4818 CRC64;

  Query Match             3.0%;  Score 217;  DB 1;  Length 1790;
  Best Local Similarity   17.5%;  Pred. No. 0.016;
  Matches  256;  Conservative  252;  Mismatches  520;  Indels  433;  Gaps  61;

Qy         36 KMVVKTFMDMDQDSEEEKELYLNLALHLASDFFLKHPGKDVRLLVACCLADIFRIYAPEA 95
              | |:|: :      ::|                    :: ||   |   :   : |:
Db        599 KSVIKSLLSFSYQIQDE-----------------DVTIKCLVTMLLGVAYEFSSKES 638

Qy         96 PYTSPDKLKDIFMFITRQL----------------------------KGLEDTKSPQ- 124
              |:     |: | |||: |                              |::|  |:
Db        639 PFPR----KEYFEFITKTLGKDNYASRIKQFKKDSYFSKVDMNEDSILTPELDETGLPKV 694

Qy        125 -FNRYFYLL--ENIAWVKS----------YNICF-ELEDSNEIFTQLYRTLFSVINNGH 169
               |: ||  |  |||  :::           | | |:|      |:|   : |:
Db        695 YFSTYFIQLFNENIYRIRTALSHDPDEEPISKISFEEVEKLQRQCTKLKGEITSL--QTE 752

Qy        170 NQKVHMHMVDLMSSIICEGDTVSQE---------------LLDTVLVNLVPAHKNLNKQ 213
                :  | :: : : : ::    |  : ::                :|:| | |:  :   :  |
Db        753 TESTHENLTEKLIALTNEHKELDEKYQILNSSHSSLKENFSILETELKNVRDSLDEMT-Q 811

Qy        214 AYDLAKALLKRTAQAIEPYITTFFNQ------------VLMLGKTS---ISDLSEHVFD 257
                |: :    |    |:  | :|  |                 :|    | :    |: : : :|
Db        812 LRDVLETKDKENQTALLEYKSTIHKQEDSIKTLEKGLETILSQKKKAEDGINKMGKDLFA 871

Qy        258 LILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPL--- 314
              |  |:  :: :            |: | |    |     :   | | :   || :|: : |: |
Db        872 LSREMQAVEEN-----CKNLQ-KEKDKSNVNHQKETKSLKEDIAAKITEIKAINENLEEM 925

Qy        315 -WQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSHDPEE-AIRHDVIV 372
               ||       :         |   |  | |:: |      :| :|    ||| ||   ::: ::      :: ::
Db        926 KIQC--NNLSKEKEHISKELVEYKSR-FQSHDNLVAKLTEKLKSLANNYKDMQAENESLI 982

Qy        373 SIVTAAKKDILLVNDHLL----------NFVRER-TLDKRWRVRKEAMMGLAQIYKKYA 420
               |   :|  :   :    :|              ||   || :::|       |: :  | |  :
Db        983 KAVEESKNESSIQLSNLQNKIDSMSQEKENFQIERGSIEKNIEQLKKTISDLEQTKEEII 1042

Qy        421 LQSAAGKDA-AKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYY 479
               :|  :  ||      ||: :|:|                                   |||
Db       1043 SKSDSSKDEYESQISLLKEK----------------------------LETA-------- 1066

Qy        480 LYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITRNLPDP 539
                 | :     |    ::|: |       | :::    |      |    |   :      |
Db       1067 --TTANDENVNKISELTK----TREELEAELAAYKNLKNELETK---------------- 1104

Qy        540 GKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLE 599
                             ||   ||   |::                   |: |   ::|    :|           |      |
Db       1105 -------------LETSEKALKEV----------KENEEHLKEEKIQLEKEATETKQQLN 1141

Qy        600 MIKFLLERIAPVHID--------TESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLE 651
                ::  || :    | |              | |:    :| |: |       |:            :::|:
```

```
Db       1142 SLRANLESLEKEHEDLAAQLKKYEEQIANKERQYNEEISQLNDEITSTQQENESIK---- 1197

Qy        652 LLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALL 711
                                         | :||   |  ::  |:|     |:  ::: : :
Db       1198 -------------------------KKNDE--LEGEVKAMKST----SEEQSNLKKSEI 1225

Qy        712 PVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIG 771
               |: : |:                  :     ||  | : | : ||::    ::
Db       1226 DALNLQIKE--------------LKKKNETNEASLLESI-KSVESETVK--------- 1259

Qy        772 HIALLAPDQFAAPWKSWVATFIVKDLL-MNDRL-----PGKKTTKLWVPDEEVSPETMVK 825
                |  |   :              |  |::  : |:|          |  :|    |::  |   |
Db       1260 -IKELQDE---------CNFKEKEVSELEDKLKASEDKNSKYLELQKESEKIKEELDAK 1308

Qy        826 IQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSA 885
                 :|: :   :   :       ||  :      ||  |            :|: |  ::   : :|:
Db       1309 TTELKIQLEKITNLSKAKEKSESELSRLKKT-------SSEERKNAEEQLEKLK------ 1355

Qy        886 IVKLAQEPCYHEI-ITLEQYQLCALAINDECYQVRQVFAQK---LHKGLSRLRLPLEYMA 941
                          :|| |    :  ::       :|:       :  |  :::|     |      | ||:     |
Db       1356 ---------NEIQIKNQAFEKERKLLNEGSSTITQEYSEKINTLEDELIRLQNENE--- 1402

Qy        942 ICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAH 1001
                  | ||: :     |:     :   :  |   :   |   ||:      |          | : :: |     :   :
Db       1403 ---LKAKE-IDNTRSELEKVSLSNDELLEE--KQNTIKS------LQDEILSYKDKITRN 1450

Qy       1002 DPDYVKVQ-----DIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGP 1056
               |    :  ::         |:| ||                 |  |  |    |:  :          |:  |  :|: ::
Db       1451 DEKLLSIERDNKRDLESLK---------EQLRAAQESKA-----KVEEGLKKLEEESSK 1495

Qy       1057 DDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEM 1116
               : |:: ||        :             :  |   |    |      |:|      :  :               :  |:          :|               ::
Db       1496 EKAEL-EKSKEMMKKLESTIESNET--ELKSSMETI------RKSDEKLEQSKKSAEEDI 1546

Qy       1117 KSFFTPGKPKTTNVLGAVN------KPLSSAGKQSQTKSSRMETVSNASSSS------NP 1164
               |:      : : :::: :|               : |  |  |               | :|||       :::                 |
Db       1547 KNL----QHEKSDLISRINESEKDIEELKSKLRIEAKSGSELETVKQELNNAQEKIRINA 1602

Qy       1165 SSPGRIKGRLDSSEMDHSENEDYTMSSP-----LPGKKSDKRDDSDLVRSELEKPRGRKK 1219
                       :| :|:  |  :  :  :       |:              |   :   :     :  |    :  :  :|          ::
Db       1603 EENTVLKSKLEDIERELKDKQAEIKSNQEEKELLTSRLKELEQELDSTQQKAQKSEEERR 1662

Qy       1220 TPVTE-QEEKLGMDDLTKLVQ---------EQKPKGSQRSRKRGHTASESDEQQWPEEKR 1269
                  | : | ||  :|:   |::                    ||   |   :  :  |:              :|  |:     ||
Db       1663 AEVRKFQVEKSQLDEKAMLLETKYNDLVNKEQAWKRDEDTVKK-----TTDSQRQEIEKL 1717

Qy       1270 LKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTM--------KTSKKGSKKK--- 1318
               ||   |:|   :||  |:          |                    :   :   |              |  :|    ||  |
Db       1718 AKE--LDNLKAENSKLKEANEDR---------SEIDDLMLLVTDLDEKNAKYRSKLKDLG 1766

Qy       1319 ----SGPPAPEEEEEEEERQSG 1335
                    |      ||::||:  :  |
Db       1767 VEISSDEEDDEEDDEEDEEEG 1787
```

```
RESULT 10
ANC1_CAEEL
ID   ANC1_CAEEL     STANDARD;      PRT;  8545 AA.
AC   Q9N4M4; O61841; O61842;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Nuclear anchorage protein 1 (Anchorage 1 protein) (Nesprin homolog).
GN   ANC-1 OR ZK973.6 OR T03A1.3/T03A1.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE OF 1-1751 FROM N.A., FUNCTION, CHARACTERIZATION, AND
RP   INTERACTION WITH F-ACTIN AND UNC-84.
RC   STRAIN=Bristol N2;
RX   MEDLINE=22264052; PubMed=12169658;
RA   Starr D.A., Han M.;
RT   "Role of ANC-1 in tethering nuclei to the actin cytoskeleton.";
RL   Science 298:406-409(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Gattung S., Goela D., Broy M.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   REVISIONS.
RA   Waterston R.;
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 689-1243 FROM N.A.
RX   MEDLINE=22296983; PubMed=12408964;
RA   Zhang Q., Ragnauth C., Greener M.J., Shanahan C.M., Roberts R.G.;
RT   "The nesprins are giant actin-binding proteins, orthologous to
RT   Drosophila melanogaster muscle protein MSP-300.";
RL   Genomics 80:473-481(2002).
CC   -!- FUNCTION: Plays a central role in nuclear and mitochondrial
CC       anchoring. Probably connects nuclei to the cytoskeleton by
CC       interacting with unc-84 at the nuclear envelope and with F-actin
CC       in the cytoplasm, creating a bridge across the nuclear envelope
CC       between the cytoskeleton and the nucleus.
CC   -!- SUBUNIT: Interacts with F-actin via its N-terminal domain.
CC       Interacts indirecty with unc-84.
CC   -!- SUBCELLULAR LOCATION: Type IV membrane protein (Potential). The
CC       largest part of the protein is cytoplasmic, while its C-terminal
CC       part is associated either with the nuclear envelope, most probably
CC       the outer nuclear membrane, or with mitochondrial membrane.
CC   -!- TISSUE SPECIFICITY: Ubiquitously expressed in all postembryonic
CC       cells.
CC   -!- DEVELOPMENTAL STAGE: First expressed in L1 larvae, and thereafter.
CC   -!- DOMAIN: The large coiled coil domains are composed of 6 nearly
CC       exact repeats of 903 residues. The last repeat is partial. These
CC       repeats are conserved in Hawai (CB4856), Australia (AB4) and
CC       Germany (RC301) strains. The length of the repeat may be
CC       maintained because of a selective advantage to keep the protein
CC       large and allow a single molecule to extent more than 0,5
CC       micrometers.
CC   -!- DOMAIN: The Klarsicht domain, which contains a potential
CC       transmembrane domain, is essential for the nuclear envelope
CC       targeting.
CC   -!- SIMILARITY: Belongs to the Nesprin family.
CC   -!- SIMILARITY: Contains 1 actin-binding domain.
CC   -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC   -!- SIMILARITY: Contains 1 Klarsicht domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AY157938; AAN35200.1; -.
DR   EMBL; AC006834; AAF40010.3; -.
DR   EMBL; AY126454; AAM95163.1; -.
DR   HSSP; P46939; 1QAG.
DR   WormPep; ZK973.6; CE33588.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR001715; Calponin-like.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF00307; CH; 2.
DR   SMART; SM00264; BAG; 6.
DR   SMART; SM00033; CH; 2.
DR   PROSITE; PS00019; ACTININ_1; 1.
DR   PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR   PROSITE; PS50021; CH; 2.
KW   Structural protein; Cytoskeleton; Actin-binding; Transmembrane;
KW   Coiled coil; Repeat.
FT   DOMAIN        1   8494       CYTOPLASMIC (POTENTIAL).
```

```
FT   TRANSMEM   8495   8513       ANCHOR FOR TYPE IV MEMBRANE PROTEIN
FT                                (POTENTIAL).
FT   DOMAIN     8514   8545       PERINUCLEAR SPACE (POTENTIAL).
FT   DOMAIN        1    325       ACTIN-BINDING.
FT   DOMAIN       23    130       CH 1.
FT   DOMAIN      222    325       CH 2.
FT   DOMAIN     3241   8199       6 X TANDEM REPEAT.
FT   REPEAT     3241   4143       1.
FT   REPEAT     4144   5097       2.
FT   REPEAT     5098   6000       3.
FT   REPEAT     6001   6903       4.
FT   REPEAT     6904   7806       5.
FT   REPEAT     7807   8199       6.
FT   DOMAIN      754    774       COILED COIL (POTENTIAL).
FT   DOMAIN     1072   1101       COILED COIL (POTENTIAL).
FT   DOMAIN     1215   1236       COILED COIL (POTENTIAL).
FT   DOMAIN     1324   1384       COILED COIL (POTENTIAL).
FT   DOMAIN     1574   1629       COILED COIL (POTENTIAL).
FT   DOMAIN     1725   1754       COILED COIL (POTENTIAL).
FT   DOMAIN     1950   1981       COILED COIL (POTENTIAL).
FT   DOMAIN     2103   2580       COILED COIL (POTENTIAL).
FT   DOMAIN     2682   2712       COILED COIL (POTENTIAL).
FT   DOMAIN     2852   2949       COILED COIL (POTENTIAL).
FT   DOMAIN     3002   3119       COILED COIL (POTENTIAL).
FT   DOMAIN     3178   3295       COILED COIL (POTENTIAL).
FT   DOMAIN     3346   3417       COILED COIL (POTENTIAL).
FT   DOMAIN     3482   3552       COILED COIL (POTENTIAL).
FT   DOMAIN     3587   3703       COILED COIL (POTENTIAL).
FT   DOMAIN     3781   3839       COILED COIL (POTENTIAL).
FT   DOMAIN     3902   4022       COILED COIL (POTENTIAL).
FT   DOMAIN     4114   4198       COILED COIL (POTENTIAL).
FT   DOMAIN     4249   4320       COILED COIL (POTENTIAL).
FT   DOMAIN     4436   4506       COILED COIL (POTENTIAL).
FT   DOMAIN     4541   4657       COILED COIL (POTENTIAL).
FT   DOMAIN     4735   4793       COILED COIL (POTENTIAL).
FT   DOMAIN     4856   4976       COILED COIL (POTENTIAL).
FT   DOMAIN     5035   5152       COILED COIL (POTENTIAL).
FT   DOMAIN     5203   5274       COILED COIL (POTENTIAL).
FT   DOMAIN     5339   5409       COILED COIL (POTENTIAL).
FT   DOMAIN     5444   5560       COILED COIL (POTENTIAL).
FT   DOMAIN     5638   5696       COILED COIL (POTENTIAL).
FT   DOMAIN     5759   5879       COILED COIL (POTENTIAL).
FT   DOMAIN     5938   6055       COILED COIL (POTENTIAL).
FT   DOMAIN     6106   6177       COILED COIL (POTENTIAL).
FT   DOMAIN     6242   6312       COILED COIL (POTENTIAL).
FT   DOMAIN     6347   6463       COILED COIL (POTENTIAL).
FT   DOMAIN     6541   6599       COILED COIL (POTENTIAL).
FT   DOMAIN     6662   6782       COILED COIL (POTENTIAL).
FT   DOMAIN     6841   6958       COILED COIL (POTENTIAL).
FT   DOMAIN     7009   7080       COILED COIL (POTENTIAL).
FT   DOMAIN     7145   7215       COILED COIL (POTENTIAL).
FT   DOMAIN     7250   7366       COILED COIL (POTENTIAL).
FT   DOMAIN     7444   7502       COILED COIL (POTENTIAL).
FT   DOMAIN     7565   7685       COILED COIL (POTENTIAL).
FT   DOMAIN     7744   7861       COILED COIL (POTENTIAL).
FT   DOMAIN     7912   7983       COILED COIL (POTENTIAL).
FT   DOMAIN     8048   8118       COILED COIL (POTENTIAL).
FT   DOMAIN     8153   8204       COILED COIL (POTENTIAL).
FT   DOMAIN     8273   8329       COILED COIL (POTENTIAL).
FT   DOMAIN     8370   8390       COILED COIL (POTENTIAL).
FT   DOMAIN     8486   8545       KLARSICHT.
FT   DOMAIN     1655   1658       POLY-ARG.
FT   DOMAIN     3021   3029       POLY-LYS.
FT   DOMAIN     3924   3932       POLY-LYS.
FT   DOMAIN     4878   4886       POLY-LYS.
FT   DOMAIN     5781   5789       POLY-LYS.
FT   DOMAIN     6684   6692       POLY-LYS.
FT   DOMAIN     7587   7595       POLY-LYS.
SQ   SEQUENCE   8545 AA;  956470 MW;  02A94D994BEE19E3 CRC64;

  Query Match             3.0%;  Score 212.5;  DB 1;  Length 8545;
  Best Local Similarity   18.4%;  Pred. No. 0.2;
  Matches  317;  Conservative  268;  Mismatches  642;  Indels  495;  Gaps   79;
```

```
Qy         19 VKEISDKIS-KEEMVRRLKMVVKTFMD----MDQDSEEEKELYLNLALHL------ASDF 67
              | ||: |:      : :: :      |      :|      : :| | :|     ||   :|        | :
Db       5221 VNEIAPKLQLVSQQLQSVPQEVPASLDEQKQLLEDVENQKHNLENLLANLPENDPTADEL 5280

Qy         68 FLKHPG-----KDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKS 122
              |         ||:   :    : |   :: |   |    :      :|   : |||:  | :| ||
Db       5281 RQKSQWDLSRLKDLLKQLGSAVGD--KLAALAAFNAARKNAEDALLDITRE-DGGDDNKS 5337

Qy        123 P---------QFNRYFYLLENIAWVKSYNICFELEDS-----NEIFTQLYRTLFSVINNG 168
              |          :        ||: :: ||         ||:|       |::   :|   |    |: |
Db       5338 PDELIDDLAKKEETVAKLLDTVSGVKP----DELDDKERAEYNDLLARL-ATAADVLKN- 5391

Qy        169 HNQKVHMHMVDLMSSIICEGDTVS-QELLDTVLVNLVPAHKNLNKQAYDLAKAL------ 221
                           :|   ::    :  |    |   : :|  ::  |||   :    ::   :: |:|:
Db       5392 -------KRAELEQAVKAKADEKSLHDSVDRIVSRLVPLVRESDELRHN-AEAVPTQYAP 5443

Qy        222 ----LKRTAQAIEPYITTF---------FNQVLMLGKTSISDLSEHVFDLILELYN---- 264
                  ||:  :|  :  |                 | :    :|  | ||  |            |:|
Db       5444 KAEELKKEVEAAKAVIANAPSSDAHVQQLEQAVATAETLIPDLEERA-----RLWNEFLA 5498

Qy        265 ----IDSHLLLSVLPQLEFKLKS------NDNEERLQVVKLL---AKMFGAKDSELASQN 311
                  ||      ::: ||:   | :              ||   |  | |:  |      ::        |:::|:
Db       5499 ARNDID-----ALIEQLQQPLDAVLAQPKRSAEEAAQDVENLRNNSQQLSDLDNKIANLQ 5553

Qy        312 K------PLWQCYLG-RF---------------NDIHVPIRLECV--KFASHCLMNHPD 346
              :        ||    |    ||                    ||:    :    |  :   : ||          |
Db       5554 RISELLDPLESAYADVRFFDVDAEQTRHQYDDVLNDVAAELEDETLLKQSASQVANEIDD 5613

Qy        347 LAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRER---TLDKRWR 403
              ::|         :  |  ||| :|      :    |  :   |   :|      |   :  |  ::|         |     |  |  :
Db       5614 ISK------MIDSTDPERSILDTIAKSDIPALKAQINRIKDRIVNADASRKHVTTDPKIA 5667

Qy        404 VRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSI-------DDRLLVE 456
                   :    :    |        |::::    |    :  |    :|     |:|          |            ||     |
Db       5668 EDLDNKLAKLQTELDDAIKTSDEHDKEQLILSLK---LNISQFEQIPLDQLKSDDLKTAE 5724

Qy        457 RIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLL------ 510
              :         :  |     |                       |   :: |  |   :   :    |      |  :
Db       5725 KEITNSLKPEEAEPL--------------LAKIQELREAKRVGDEARSAAHDQIVALEKE 5770

Qy        511 --------------DLIKQPK---TDASVKAIFSKVMV-------ITRNLPDPGKAQDF 545
                                       |    |  |:      :  |  |  :  :|  ::                   |||        |
Db       5771 AEDVTAKESAKKKKDKKKSPQEMIDELSAKVVEAKALIPKIEEAAKNENLP-----ADD 5825

Qy        546 MKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAE----GCVREITKKLGNPKQ-------PT 594
              |   |:: :  |      |   :|   ||       || |                  |  :|:||:             |:
Db       5826 KPKAEQLVSNLEAFVKDVETQVSE----KQDELDKLNNANDAIKRLGDALDDAEKTVVPS 5881

Qy        595 N-PFLEMIKFLLERIAPVHIDT-------------------ESISALIKQVNKSIDGTA 633
              : |  |     |    :||||  |: |                                   :    :    :    :  |:|   |
Db       5882 SVPALSEFK---DRIAP-HLATLVEAVNDVPASVEPSAVALRDRAAKFVSDLEKNIQKTG 5937

Qy        634 DDE------------------------DEGVPTDQA------IRAGLELLKVLSFTHPI 662
              |||                                         ::   |  |  |          ::|  :|   |   |:
Db       5938 DDEKRADELKNDVGNAVKNVEDVVSKYQNQPQPLDVAKDDANKLKATVEQLTKLA----- 5992

Qy        663 SFHSAETFESLLACLKMDDEKVAEAALQ----------IFKNTGSKIEEDFPHIRSALLP 712
                 |::   :    :|      |   :    |:    ||                    :       ::|  :      ::       |
Db       5993 --ESSDKIDPQVAKDIKDSKTKAKELLQALEKAIPQEDAIRREQAEINDRLNNLEKELTK 6050

Qy        713 VLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIGH 772
              |     |  :                 |:    :    :  ::     |              |:     ||  |     :    |     |
Db       6051 VDEFKPED--------ALPIVDQLAANTNTL---------KTATDSNNEKAVAPSSLISH 6093

Qy        773 IALLAPDQFAAPWKSWVATFIVKD--------LLMNDRLPGKKTTKLWVPDEE---VSPE 821
              |:              |  |  :         |                 :|:        ||  :     ::       |    |
Db       6094 DDLVV----GLPEKVFQLQHAIDDKKQALNKAAAVNEIAP-----KLQLVSQQLQSVPQE 6144

Qy        822 TMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLR-- 879
                  :     |  ::    :     |:|                 ||    :     :|     |    :  |:   |:|||:
Db       6145 VPASLDEQKQLLEDVENQKHNLE-------NLLANLPENDPTADELRQKSQWDLSRLKDL 6197
```

```
Qy        880 --------------LAAGSAIVKLAQEPCY--------------HEIITLEQYQLCALAI 911
                            ||| :|  | |::                  |:|
Db       6198 LKQLGSAVGDKLAALAAFNAARKNAEDALLDITREDGGDDNKSPDELI------------ 6245

Qy        912 NDECYQVRQVFAQKLH--KGLSRLRLP----LEYMAICA--LCAKDPVKERRAHARQCLV 963
               |:  :  :  |: |    |:    |      ||  : |    | | :| :||   |
Db       6246 -DDLAKKEETVAKLLDTVSGVKPDELDDKERAEYNDLLARLATAADVLKNKRAELEQA-- 6302

Qy        964 KNINVRREYLKQHAAVSEKLLSLLP-----------------EYV---------VPYTIH 997
                :  : :    | :|   :  |:|                 :|          |
Db       6303 --VKAKADEKSLHDSVDRIVSRLVPLVRESDELRHNAEAVPTQYAPKAEELKKEVEAAKA 6360

Qy        998 LLAHDPDY-VKVQDIEQ--------LKDVKECLWFVLEILMAKNENNSHAFIRKMVENIK 1048
              ::|: |     || :||        : |::|      | | |:|:      |  ::|::
Db       6361 VIANAPSSDAHVQQLEQAVATAETLIPDLEERARLWNEFLAARND------IDALIEQLQ 6414

Qy       1049 QTKDA-----------QGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARF 1097
              |  ||              |   :      : |:   |   :  :  |:  :      ||
Db       6415 QPLDAVLAQPKRSAEEAAQDVENLRNNSQQLSDLDNKIANLQRISELLDPLESAYADVRF 6474

Qy       1098 F----TQPDKNFSNTKNYLPPEM--KSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSR 1151
              |     |    : :  | :  |:  ::       :  | :  ::| : |    :  : |
Db       6475 FDVDAEQTRHQYDDVLNDVAAELEDETLLKQSASQVANEIDDISKMIDS----TDPERSI 6530

Qy       1152 METVSNASSSSNPSSPGRIKGRL------------DSSEMDHSENEDYTMSSPLPG--KK 1197
              ::|:: :   :  :   ||| |:            |    :  :|:   : : |    |
Db       6531 LDTIAKSDIPALKAQINRIKDRIVNADASRKHVTTDPKIAEDLDNKLAKLQTELDDAIKT 6590

Qy       1198 SDKRDDSDLVRSELEKPRGRKKTPVTEQE----EKLGMDDL----TKLVQEQKPKGSQ-- 1247
              ||: |   |: |         |  ::: |    ::|  |||     ::    ||: ::
Db       6591 SDEHDKEQLILS--------LKLNISQFEQIPLDQLKSDDLKTAEKEITNSLKPEEAEPL 6642

Qy       1248 -------RSRKR-GHTASESDEQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGG 1299
                     |  || |  |  :   |    ::  ||:   |  :    || |:  | : :
Db       6643 LAKIQELREAKRVGDEARSAAHDQIVALEKEAEDVTAKESAKKK--KKDKKKSPQEMIDE 6700

Qy       1300 GTPK--EEPTMKTSKKGSKKKSGPPAPEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAE 1357
               : |  |   :    : : |    || :: : |:  | | |   |  : :|| :  :  :
Db       6701 LSAKVVEAKALIPKIEEAAKNENLPADDKPKAEQLVS-NLEAFVKDVETQVSEKQDELDK 6759

Qy       1358 SPESS--------AIESTQSTPQKGRGRPSKTPSPSQPKKNV 1391
                 ::        |::  : |       ||  |: |: |  :
Db       6760 LNNANDAIKRLGDALDDAEKTVV-----PSSVPALSEFKDRI 6796


RESULT 11
NFH_RAT
ID   NFH_RAT        STANDARD;      PRT;    831 AA.
AC   P16884; Q63368;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neurofilament triplet H protein (200 kDa neurofilament protein)
DE   (Neurofilament heavy polypeptide) (NF-H) (Fragment).
GN   NEFH OR NFH.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=89065087; PubMed=3143606;
RA   Breen K.C., Robinson P.A., Wion D., Anderton B.H.;
RT   "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).
RT   Identification of putative phosphorylation sites.";
RL   FEBS Lett. 241:213-218(1988).
RN   [2]
RP   SEQUENCE OF 37-831 FROM N.A.
RX   MEDLINE=88309090; PubMed=2457365;
RA   Dautigny A., Pham-Dinh D., Roussel C., Felix J.M., Nussbaum J.L.,
RA   Jolles P.;
RT   "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and
RT   in situ detection.";
RL   Biochem. Biophys. Res. Commun. 154:1099-1106(1988).
RN   [3]
RP   SEQUENCE OF 1-89 AND 243-313 FROM N.A.
RX   MEDLINE=87080760; PubMed=2878828;
RA   Robinson P.A., Wion D., Anderton B.H.;
RT   "Isolation of a cDNA for the rat heavy neurofilament polypeptide
RT   (NF-H).";
RL   FEBS Lett. 209:203-205(1986).
RN   [4]
RP   SEQUENCE OF 318-831 FROM N.A.
RX   MEDLINE=89184647; PubMed=2928342;
RA   Lieberburg I., Spinner N., Snyder S., Anderson J., Goldgaber D.,
RA   Smulowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;
RT   "Cloning of a cDNA encoding the rat high molecular weight
RT   neurofilament peptide (NF-H): developmental and tissue expression in
RT   the rat, and mapping of its human homologue to chromosomes 1 and
RT   22.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
CC   -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC       AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC       NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC       SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC   -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC       PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC       THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC       INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC       OF AXONAL CALIBER.
CC   -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC       OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC       LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC       COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
CC       ONWARD AND IS LONGER DUE TO A FRAMESHIFT.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M37227; AAA41693.1; ALT_FRAME.
DR   EMBL; X13804; CAA32038.1; ALT_FRAME.
DR   EMBL; M21964; AAA41695.1; -.
DR   EMBL; J04517; AAA41692.1; -.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.
FT   NON_TER       1      1
FT   DOMAIN      276    641       51 X 3 AA TANDEM REPEATS OF K-S-P.
FT   CONFLICT    164    164       L -> I (IN REF. 2).
FT   CONFLICT    185    185       I -> S (IN REF. 2).
FT   CONFLICT    193    193       L -> T (IN REF. 2).
FT   CONFLICT    199    199       M -> T (IN REF. 2).
FT   CONFLICT    346    346       K -> N (IN REF. 1).
FT   CONFLICT    373    373       A -> V (IN REF. 1 AND 4).
FT   CONFLICT    482    482       G -> E (IN REF. 2 AND 4).
FT   CONFLICT    485    485       P -> S (IN REF. 2).
FT   CONFLICT    570    571       RK -> KE (IN REF. 2 AND 4).
FT   CONFLICT    591    591       P -> T (IN REF. 2 AND 4).
FT   CONFLICT    727    727       A -> V (IN REF. 4).
FT   CONFLICT    757    759       AAP -> GST (IN REF. 4).
FT   CONFLICT    769    769       T -> L (IN REF. 2).
FT   CONFLICT    775    775       R -> P (IN REF. 2 AND 4).
SQ   SEQUENCE   831 AA;  89486 MW;  1B0973C3F13EF768 CRC64;

  Query Match             2.9%;  Score 212;  DB 1;  Length 831;
  Best Local Similarity   22.9%;  Pred. No. 0.01;
  Matches   85;  Conservative   56;  Mismatches  150;  Indels   80;  Gaps   13;
```

```
Qy       1048 KQTKDAQGPDDAKMNEKLYT-VCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFS 1106
              |    :|: | : |  ||  : | : | ::  :||   :     |: : |       |:|
Db        508 KSPAEAKSPVEVKSPEKAKSPVKEGAKSLAEAKSPEKAKSPVKEEIKPPAEVKSPEK--- 564

Qy       1107 NTKNYLPPEMKSFFTPGKPKTTNVLG--------------------AVNKPLSSAGKQS 1145
               |: :  | ||   | | || :|                         |   |    |
Db        565 -AKSPMRKEAKS---PEKAKTLDVKSPEAKPPAKEEAKRPADIRSPEQVKSPAKEEAKSP 620

Qy       1146 QTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKRDDSD 1205
               : : :| | |  |          ||      : : |      |: | ::|   |  |:|::
Db        621 EKEETRTEKV--APKKEEVKSPVE---EVKAKEPPKKVEEEKTPATPKTEVKESKKDEAP 675

Qy       1206 LVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESDEQQWP 1265
                   | :||:  :| |:||                     ||| |      |:
Db        676 ---KEAQKPKAEEKEPLTE----------------KPKDSPGEAKK------------ 702

Qy       1266 EEKRLKEDILENEDEQNSPPKKG-KRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAP 1324
              || : |:          |: :| | | |    ||          :||:  : |:  ||:  | ||
Db        703 EEAKEKK---AAAPEEETPAKLGVKEEAKPKEKAEDAKAKEPSKPSEKEKPKKEEVPAAP 759

Qy       1325 EEEEEEERQSGNTEQKSKSKQHRVSRRAQ-------QRAESPESSAIESTQSTPQKGRGR 1377
              |::: :| ::  || | : :: ::  :|:         |     |::    | : || ||
Db        760 EKKDTKEEKT--TESKKREEKPKMEAKAKEEDKGLPQEPSKPKTEKAEKSSSTDQKDSQP 817

Qy       1378 PSKTPSPSQPK 1388
                 | |    |
Db        818 SEKAPEDKAAK 828


RESULT 12
MDN1_HUMAN
ID   MDN1_HUMAN     STANDARD;      PRT;  5596 AA.
AC   Q9NU22; O15019;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Midasin (MIDAS-containing protein).
GN   MDN1 OR KIAA0301.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   PubMed=12102729;
RA   Garbarino J.E., Gibbons I.R.;
RT   "Expression and genomic analysis of midasin, a novel and highly
RT   conserved AAA protein distantly related to dynein.";
RL   BMC Genomics 3:18-18(2002).
RN   [2]
RP   SEQUENCE OF 1255-2356 AND 3550-5596 FROM N.A.
RC   TISSUE=Testis;
RA   Tracey A.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 3550-5596 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97349984; PubMed=9205841;
RA   Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. VII.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 4:141-150(1997).
CC   -!- FUNCTION: May function as a nuclear chaperone and be involved in
CC       the assembly/disassembly of macromolecular complexes in the
CC       nucleus.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: Contains 1 VWFA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF503925; AAM77722.1; -.
DR   EMBL; AL096678; CAB86660.1; -.
DR   EMBL; AL096678; CAB86661.1; -.
DR   EMBL; AB002299; BAA20761.1; -.
DR   Genew; HGNC:18302; MDN1.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0016887; F:ATPase activity; NAS.
DR   GO; GO:0003754; F:chaperone activity; NAS.
DR   GO; GO:0006461; P:protein complex assembly; NAS.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR002035; VWF_A.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 7.
DR   SMART; SM00327; VWA; 1.
DR   PROSITE; PS50234; VWFA; 1.
KW   Chaperone; ATP-binding; Repeat; Nuclear protein.
FT   NP_BIND     329    336       ATP (POTENTIAL).
FT   NP_BIND     677    684       ATP (POTENTIAL).
FT   NP_BIND    1084   1091       ATP (POTENTIAL).
FT   NP_BIND    1390   1397       ATP (POTENTIAL).
FT   NP_BIND    1753   1760       ATP (POTENTIAL).
FT   NP_BIND    2066   2073       ATP (POTENTIAL).
FT   DOMAIN     3566   3573       POLY-GLU.
FT   DOMAIN     4784   4791       POLY-GLU.
FT   DOMAIN     5008   5013       POLY-GLU.
FT   DOMAIN     5182   5187       POLY-GLU.
FT   DOMAIN     5384   5583       VWFA.
FT   CONFLICT   2287   2312       RLFLSMDPVHGDISRAMRNRGLEIYI -> S (IN REF.
FT                                2).
SQ   SEQUENCE   5596 AA;  632802 MW;  586C62616A1F96D4 CRC64;

  Query Match              2.9%;  Score 210.5;  DB 1;  Length 5596;
  Best Local Similarity   18.4%;  Pred. No. 0.14;
  Matches  290;  Conservative  243;  Mismatches  530;  Indels  511;  Gaps   76;

Qy         62 HLASDFFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLE--- 118
              ||  | : :|| :  | :|      |  | :  | :||           |:: | |||
Db       3740 HLLQD-WPEHPALEQLLVV----MDRIRSF----PLSSP---------ISKFLNGLEILL 3781

Qy        119 ---------DTKSPQFNRYFYLLEN--IAWVKSYNICFELEDSN------EIFTQLYRTL 161
                       :::    ::  |:    | | |    |: :   |       |  |: : ::
Db       3782 AKAQDWEENASRALSLRKHLDLISQMIIRWRKLELNCWSMSLDNTMKRHTEKSTKHWFSI 3841

Qy        162 FSVINNGHNQK--------VHMHMVDLMSSI--ICEGDTVSQ----------------- 193
              : ::  | |:               | :: |:|::   || :: :
Db       3842 YQMLEK-HMQEQTEEQEDDKQMTLMLLVSTLQAFIEGSSLGEFHVRLQMLLVFHCHVLLM 3900

Qy        194 ------ELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTF-----FNQVLM 242
                    : | :|| ||     :  || :| :| :      :|  :  |     :| |
Db       3901 PQVEGKDSLCSVLWNLY----HYYKQFFDRVQAKIVELRSPLEKELKEFVKISKWNDVSF 3956

Qy        243 LG-KTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMFG 301
               | |:           :|  : :   |     |       |  :| ||:        :
Db       3957 WSIKQSVEKTHRTLFKFMKKFEAVLSEPCRS-------SLVESDKEEQ---PDFLPRPTD 4006

Qy        302 AKDSELAS-QNK------------------PLWQC-----------YLGRFNDIHVPIR 330
                |||:| ||                    | | |           | |      :|
Db       4007 GAASELSSIQNLNRALRETLLAQPAAGQATIPEW-CQGAAPSGLEGELLRRLPKLRKRMR 4065

Qy        331 LECVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLL 390
                |: |       :  | | : | ::      :  :     ::        : .:  :|:|:
Db       4066 KMCLTFMKESPL--PRLVEGLDQF----TGEVISSVSELQSLKVEPSAEKE--------- 4110

Qy        391 NFVRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDK----LLHIY-- 444
                ::|:  |    ::|:     |: ::|  |       |      | :|| : |    :||::
Db       4111 ---KQRSEAKHILMQKQ--RALSDLFKHLA---KIGLSYRKGLAWARSKNPQEMLHLHPL 4162
```

```
Qy        445 -YQNSI---------DDRLLVERIFAQYMVPHNLETTERMKCLYYLYATL----DLNAVK 490
              |:::           | ||| |                ::     |  | |  :|      |||
Db       4163 DLQSALSIVSSTQEADSRLLTE------------ISSSWDGCQKYFYRSLARHARLNAAL 4210

Qy        491 A-------LNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITRNL------- 536
              |       :  : :|:     | :|    |::| ::  ::    |:  :| |||
Db       4211 ATPAKEMGMGNVERCRGFSAHLMK---MLVRQRRSLTTL----SEQWIILRNLLSCVQEI 4263

Qy        537 ---------PDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVS---------PTCSCKQAE 577
                       |     || ::::|: |    :  :   |  ::|:           |:
Db       4264 HSRLMGPQAYPVAFPPQDGVQQWTERL---QHLAMQCQILLEQLSWLLQCCPSVGPAPGH 4320

Qy        578 GCVREITKKLGNPKQP-----------TNPFLEMIKFLLERIAPV------------- 611
              | |    : || |  |                |::|    |     :|:
Db       4321 GNV----QVLGQPPGPCLEGPELSKGQLCGVVLDLIPSNLSYPSPIPGSQLPSGCRMRKQ 4376

Qy        612 -HIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETF 670
               |:  :| : |                      | :  |  : ::| ::  |:   :     ||| : |
Db       4377 DHLWQQSTTRL--------------TEMLKTIKTVKADVD--KIRQQSCETLFHSWKDF 4419

Qy        671 E---SLLACL---------------------KMDDEKVAEAALQIFKNTGSKIEEDFPH 705
              |    |  | |:||                      : |  :      |:  :   ||   ||
Db       4420 EVCSSALSCLSQVSVHLQGLESLFILPGMEVEQRDSQMALVESLEYVRGEISKAMADFTT 4479

Qy        706 IRSALL--------PVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDP 757
              :: ||          :|     :     | : || |      : | :  : |       |
Db       4480 WKTHLLTSDSQGGNQMLDEGFVEDFSEQMEIAIRAILCAIQNLEERKNEKAEENTDQASP 4539

Qy        758 ----SNLEHLITPLVTIGHIALLAPDQFAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWV 813
                  :  | | :       ||:  |  | |                              |
Db       4540 QEDYAGFERLQS-----GHLTKLLEDDF------------------------------WA 4564

Qy        814 PDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRL-----------LTTILHSDG 862
                 :||    : :| |   :  ||    :: : ||:   |           |  :| |
Db       4565 ---DVS---TLHVQKIISAISELLERLKSYGEDGTAAKHLFFSQSCSLLVRLVPVLSSYS 4618

Qy        863 DLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVF 922
              ||              : ||   :  ||                     |  ||          |||
Db       4619 DLV-------LFFLTMSLATHRSTAKL----------------LSVLA---------QVF 4646

Qy        923 AQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEK 982
                :    ||      || |:|          :|    |
Db       4647 TELAQKGFC---LPKEFM-------EDSAGEGATE------------------------ 4671

Qy        983 LLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRK 1042
                                    || :   :  :  | :||| :              ||          :|
Db       4672 ----------------FHDYEGGGIGEGEGMKDVSD---------QIGNEEQVEDTFQK 4705

Qy       1043 MVENIKQTKDA----QGPDDA-KMNE----KLYTVCDVAMNIIMSKSTTY---------S 1084
                 |  |:    | |:  |:|  :|:|                  || :
Db       4706 GQEKDKEDPDSKSDIKGEDNAIEMSEDFDGKMHDGELEEQEEDDEKSDSEGGDLDKHMGD 4765

Qy       1085 LESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFT-PGKPKTTNVLGAVNKPLSSAGK 1143
              | · :     |  |  :     |:       :        : |:  | ||    :  |  |  :   |   |
Db       4766 LNGEEADKLDERLWGDDDEEEDEEEE----DNKTEETGPGMDEEDSELVAKDDNLDS-GN 4820

Qy       1144 QSQTKS--SRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSS---PLPGKKS 1198
               :: ||    : |      | :              :|   ::|   : | :| : |   :       | | :
Db       4821 SNKDKSQQDKKEEKEEAEADDGGQGEDKINEQIDERDYDENEVDPYHGNQEKVPEP-EAL 4879

Qy       1199 DKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHT--- 1255
              |  || :|    : |     |  : |     | ||:  ::         ::|:  :       : :|| |
Db       4880 DLPDDLNL---DSEDKNGGEDTDNEEGEEENPLE-----IKEKPEEAGHEAEERGETETD 4931

Qy       1256 ASESDEQQWPEEKRLKEDILENEDEQN--SPPKKGKRGRPPKPLGGGTPKEEPTMKTSKK 1313
               :||   |  |||    :    :|   |   |:|  :    :   :    | :     :  : ::: :::   |
Db       4932 QNESQSPQEPEEGPSEDDKAEGEEEMDTGADDQDGDAAQHPEE---HSEEQQQSVEEKDK 4988

Qy       1314 GSKKKSGPPAP-----EEEEEEERQSGNTEQK----SKSKQHRVSRRAQQRAESPESS-A 1363
               :  :: |      |          :  :|||||: :||::           :    |       |: ::: |
Db       4989 EADEEGGENGPADQGFQPQEEEEREDSDTEEQVPEALERKEH--ASCGQTGVENMQNTQA 5046

Qy       1364 IESTQSTPQKGRGR 1377
                  :|   :  |:| :|:
Db       5047 MELAGAAPEKEQGK 5060
```

```
RESULT 13
BPA1_MOUSE
ID   BPA1_MOUSE     STANDARD;      PRT;  7389 AA.
AC   Q91ZU6; Q91ZU7;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Bullous pemphigoid antigen 1, isoforms 1/2/3/4 (BPA) (Hemidesmosomal
DE   plaque protein) (Dystonia musculorum protein) (Dystonin).
GN   BPAG1 OR DST.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND TISSUE SPECIFICITY.
RC   STRAIN=BALB/c; TISSUE=Muscle, and Neuron;
RX   MEDLINE=21405767; PubMed=11514586;
RA   Leung C.L., Zheng M., Prater S.M., Liem R.K.H.;
RT   "The BPAG1 locus: alternative splicing produces multiple isoforms with
RT   distinct cytoskeletal linker domains, including predominant isoforms
RT   in neurons and muscles.";
RL   J. Cell Biol. 154:691-697(2001).
RN   [2]
RP   SEQUENCE OF 6693-7389 FROM N.A. (ISOFORMS 3 AND 4).
RC   STRAIN=C57BL/6J; TISSUE=Fetal skin, and Fetal spinal cord;
RX   MEDLINE=22354683; PubMed=12466851;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
CC   -!- FUNCTION: Cytoskeletal linker protein. Anchors keratin-containing
CC       intermediate filaments to the inner plaque of hemidesmosomes. The
CC       proteins may self-aggregate to form filaments or a two-dimensional
CC       mesh (By similarity).
CC   -!- SUBUNIT: Homodimer. Interacts with the neuronal intermediate
CC       filament protein, Prph (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=7;
CC       Name=2; Synonyms=b;
CC         IsoId=Q91ZU6-1; Sequence=Displayed;
CC       Name=1; Synonyms=a;
```

```
CC         IsoId=Q91ZU6-2; Sequence=VSP_050483;
CC       Name=3;
CC         IsoId=Q91ZU6-3; Sequence=VSP_050484, VSP_050485, VSP_050486;
CC         Note=No experimental confirmation available;
CC       Name=4;
CC         IsoId=Q91ZU6-4; Sequence=VSP_050485, VSP_050486;
CC         Note=No experimental confirmation available;
CC       Name=5; Synonyms=e;
CC         IsoId=Q91ZU8-1; Sequence=External;
CC       Name=6; Synonyms=n1;
CC         IsoId=Q60824-1; Sequence=External;
CC       Name=7; Synonyms=n2;
CC         IsoId=Q60824-2; Sequence=External;
CC   -!- TISSUE SPECIFICITY: Expressed at high levels in the heart and
CC       skeletal muscle and at low levels in the skin in the adult.
CC       Expressed in the myocardium, skeletal muscle masses, vertebrae
CC       cartilage, and epithelia of the tongue of 14.5 day embryos.
CC   -!- SIMILARITY: Belongs to the plakin or cytolinker family.
CC   -!- SIMILARITY: Contains 1 actin-binding domain.
CC   -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC   -!- SIMILARITY: Contains 1 SH3 domain.
CC   -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC   -!- SIMILARITY: Contains 9 plectin repeats.
CC   -!- SIMILARITY: Contains 27 spectrin repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF396879; AAK83384.1; -.
DR   EMBL; AF396878; AAK83383.1; -.
DR   EMBL; AK051626; BAC34695.1; -.
DR   EMBL; AK037206; BAC29753.1; -.
DR   MGD; MGI:104627; Dst.
DR   GO; GO:0005737; C:cytoplasm; ISS.
DR   GO; GO:0030056; C:hemidesmosome; IDA.
DR   GO; GO:0005200; F:structural constituent of cytoskeleton; ISS.
DR   GO; GO:0045104; P:intermediate filament cytoskeleton organiza. . .; ISS.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR001715; Calponin-like.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR003108; GAS2.
DR   InterPro; IPR001101; Plectin_repeat.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF00307; CH; 2.
DR   Pfam; PF00036; efhand; 2.
DR   Pfam; PF02187; GAS2; 1.
DR   Pfam; PF00681; Plectin; 5.
DR   Pfam; PF00018; SH3; 1.
DR   Pfam; PF00435; spectrin; 26.
DR   ProDom; PD000012; EF-hand; 1.
DR   SMART; SM00033; CH; 2.
DR   SMART; SM00054; EFh; 2.
DR   SMART; SM00243; GAS2; 1.
DR   SMART; SM00250; PLEC; 9.
DR   SMART; SM00150; SPEC; 34.
DR   PROSITE; PS00019; ACTININ_1; 1.
DR   PROSITE; PS50021; CH; 2.
DR   PROSITE; PS00018; EF_HAND; 2.
KW   Actin-binding; Coiled coil; Repeat; SH3 domain; Structural protein;
KW   Cytoskeleton; Cell adhesion; Calcium; Calcium-binding;
KW   Alternative splicing.
FT   DOMAIN       31    255       ACTIN-BINDING.
FT   DOMAIN       35    138       CH 1.
FT   DOMAIN      151    252       CH 2.
FT   REPEAT      590    667       SPECTRIN 1.
FT   REPEAT      675    770       SPECTRIN 2.
FT   DOMAIN      889    941       SH3.
FT   REPEAT     1260   1356       SPECTRIN 3.
FT   REPEAT     1537   1581       PLECTIN 1.
FT   REPEAT     1582   1619       PLECTIN 2.
FT   REPEAT     1657   1694       PLECTIN 3.
FT   REPEAT     1695   1732       PLECTIN 4.
FT   REPEAT     1735   1770       PLECTIN 5.
FT   REPEAT     1771   1808       PLECTIN 6.
FT   REPEAT     1811   1846       PLECTIN 7.
FT   REPEAT     1847   1884       PLECTIN 8.
FT   REPEAT     1886   1922       PLECTIN 9.
FT   REPEAT     3814   3914       SPECTRIN 4.
FT   REPEAT     4053   4152       SPECTRIN 5.
FT   REPEAT     4270   4346       SPECTRIN 6.
FT   REPEAT     4409   4517       SPECTRIN 7.
FT   REPEAT     4522   4620       SPECTRIN 8.
FT   REPEAT     4623   4729       SPECTRIN 9.
FT   REPEAT     4742   4840       SPECTRIN 10.
FT   REPEAT     4851   4949       SPECTRIN 11.
FT   REPEAT     5177   5278       SPECTRIN 12.
FT   REPEAT     5288   5385       SPECTRIN 13.
FT   REPEAT     5397   5497       SPECTRIN 14.
FT   REPEAT     5506   5605       SPECTRIN 15.
FT   REPEAT     5646   5714       SPECTRIN 16.
FT   REPEAT     5725   5824       SPECTRIN 17.
FT   REPEAT     5946   6046       SPECTRIN 18.
FT   REPEAT     6055   6155       SPECTRIN 19.
FT   REPEAT     6165   6265       SPECTRIN 20.
FT   REPEAT     6274   6372       SPECTRIN 21.
FT   REPEAT     6383   6480       SPECTRIN 22.
FT   REPEAT     6492   6592       SPECTRIN 23.
FT   REPEAT     6602   6701       SPECTRIN 24.
FT   REPEAT     6710   6808       SPECTRIN 25.
FT   REPEAT     6826   6914       SPECTRIN 26.
FT   REPEAT     6962   7020       SPECTRIN 27.
FT   CA_BIND    7028   7040       EF-HAND 1 (POTENTIAL).
FT   CA_BIND    7064   7076       EF-HAND 2 (POTENTIAL).
FT   DOMAIN      452    486       COILED COIL (POTENTIAL).
FT   DOMAIN      638    703       COILED COIL (POTENTIAL).
FT   DOMAIN      737    773       COILED COIL (POTENTIAL).
FT   DOMAIN     1003   1138       COILED COIL (POTENTIAL).
FT   DOMAIN     1195   1247       COILED COIL (POTENTIAL).
FT   DOMAIN     1413   1455       COILED COIL (POTENTIAL).
FT   DOMAIN     1504   1527       COILED COIL (POTENTIAL).
FT   DOMAIN     3336   3359       COILED COIL (POTENTIAL).
FT   DOMAIN     3539   3715       COILED COIL (POTENTIAL).
FT   DOMAIN     3809   3893       COILED COIL (POTENTIAL).
FT   DOMAIN     3957   3978       COILED COIL (POTENTIAL).
FT   DOMAIN     4006   4043       COILED COIL (POTENTIAL).
FT   DOMAIN     4159   4209       COILED COIL (POTENTIAL).
FT   DOMAIN     4270   4311       COILED COIL (POTENTIAL).
FT   DOMAIN     4426   4563       COILED COIL (POTENTIAL).
FT   DOMAIN     4696   4735       COILED COIL (POTENTIAL).
FT   DOMAIN     4847   5097       COILED COIL (POTENTIAL).
FT   DOMAIN     5173   5233       COILED COIL (POTENTIAL).
FT   DOMAIN     5570   5603       COILED COIL (POTENTIAL).
FT   DOMAIN     5717   5739       COILED COIL (POTENTIAL).
FT   DOMAIN     5787   5809       COILED COIL (POTENTIAL).
FT   DOMAIN     6010   6089       COILED COIL (POTENTIAL).
FT   DOMAIN     6116   6164       COILED COIL (POTENTIAL).
FT   DOMAIN     6277   6318       COILED COIL (POTENTIAL).
FT   DOMAIN     6381   6417       COILED COIL (POTENTIAL).

  Query Match             2.9%;  Score 209;  DB 1;  Length 7389;
  Best Local Similarity   18.2%;  Pred. No. 0.24;
  Matches  320;  Conservative  247;  Mismatches  591;  Indels  600;  Gaps   77;

Qy         17 PGVKE-----------ISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLALHLAS 65
              | :||            : ::  | |: |:|: ::|:| |::| :|        |   || |
Db       4717 PKIKEAQEVCEDLSALVKEEYLKAELSRQLEGILKSFKDIEQKTE-------NHVQHLQS 4769

Qy         66 DFFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQF 125
                                 ||  :  |:              ||   ::    |  |   ||
Db       4770 ---------------ACASSHQFQ-----------QMSKDFQAWL--DAKKEEQRDSPPI 4801
```

```
Qy        126 NRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSSII 185
              :    :||         ::     :|  :  :||:          |:  |  |          ::
Db       4802 SAKLDVLE--------SLLNSQKDFGKTFTEQSNIYEKTISEGEN------------LL 4840

Qy        186 CEGDTVSQELLDTVLVNLVPAHKNLNKQAYD--------LAKAL-LKRTAQAIEPYIT-- 234
               :     :    |    |   :          ||  :       | |||  :    :  : |:|
Db       4841 LKTQGAEKAALQLQLNTMKTDWDRFRKQVKEREEKLKDSLEKALKYREQVETLRPWIDRC 4900

Qy        235 ------TFFNQVLMLGKTSISDLS------EHVFDLILELYNIDSHLLLSVLPQLEFKLK 282
                     |:       ::||::|         :|  |  :  |||  |   ::  ||||   ::: :
Db       4901 QHSLDGVTFSLDPTESESSIAELKSLQKEMDHHFGM-LELLNNTANSLLSVC-EVDKEAV 4958

Qy        283 SNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCY---------------------LG 320
              : :|:   :: |   :  :      :|    |  :   :    :                   ||
Db       4959 TEENQSLMEKVNRVTEQLQSKTVSLENMAQKFKEFQEVSRDTQRQLQDTKEQLEVHHSLG 5018

Qy        321 --RFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAA 378
                ::: |:  :  |:  :  :        |      | ||  |  :  |  |   |  :  :          ||::   | |
Db       5019 PQAYSNKHLSV-LQAQQKSLQTLKQQVDEAKRLAQDLVVEAADSKGT--SDVLLQAETLA 5075

Qy        379 KKDILLVND---------------HLLNFVRE---------------------RTLDK 400
              ::     |                       |   |  :||                               || |
Db       5076 EEHSELSQQVDEKCSFLETKLQGLGHFQNTIREMFSQFTECDDELDGMAPVGRDAETLRK 5135

Qy        401 R------WRVRKEAMM----------------------GLAQIYKKYALQSAAGKDAA 430
              :        :   :  ||:|                              |:  :   :   :   |      |  |
Db       5136 QKACMQTFLKKLEALMASNDSANRTCKMMLATEETSPDLIGVKRDLEALSKQCNKLLDRA 5195

Qy        431 K----QIAWIKDKLLHIYYQ-----------------------NSIDDRLLVERIF-- 459
              |       |:       :||       :  :                          :|:  :|  |  ::|
Db       5196 KTREEQVDGATEKLEEFHRKLEEFSTLLQKAEEHEESQGPVGTETETINQQLDVFKVFQK 5255

Qy        460 -----------AQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWK---------- 497
                          ::          :::           |     |        :::|   ||
Db       5256 EEIEPLQVKQQDVNWLGQGLIQSAAANTCTQGLEHDLD-----SVNSRWKTLNKKVAQRT 5310

Qy        498 --CQNLLRH--QVKDLLDLI-------------KQPKTDAS-VKAIFSKVMVITRNLPD 538
                  |   |   |    :  :|  |:  :                      |  |       |||     :    ::  |  |  |
Db       5311 SQLQEALLHCGRFQDALESLLSWMADTEELVANQKPPSAEFKVVKAQIQEQKLLQRLLED 5370

Qy        539 PGKAQDFMKKFTQVL------EDDEKIRKQLEVLVS--------PTCSCKQAEGCVREIT 584
                   :  :|:   :   :           |   |:  :||  :|  |                      :|  ||  :   :
Db       5371 RKSTVEVIKREGEKIAASAEPADRVKLTRQLSLLDSRWEALLSRAEARNRQLEG-ISVVA 5429

Qy        585 KKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQVNKSIDGTADDED---EGVP 641
              ::         :|  |  :|   ::   |        |:         :       |  |  :|::            ||
Db       5430 QEFHGTLEPLNEWLTAVEKKLANSEPIGTQAPKLEEQISQ-HKAL-----QEDILLRKQS 5483

Qy        642 TDQAIRAGLELLK------VLSFTH------------PISFHSAETFESLL-------- 674
                |||:   ||||||        ||                           :|        |:|  |  |
Db       5484 VDQALLNGLELLKQTTGDEVLIIQDKLEAIKARYKDITKLSADVAKTLEHALQLAGQLQS 5543

Qy        675 ----ACLKMDDEKV-----------AEAALQI------FKN---TGSKIEEDFPHIRSAL 710
                      |   :|   :|                   |||  :|              ||      :        :  :  |||
Db       5544 MHKELCNWLDKVEVELLSYETQGLKGEAASQVQERQKELKNEVRSNKALVDSLNEVSSAL 5603

Qy        711 LPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTI 770
              | ::    ::|:|                         |     |       |||      :|:       |   :     :
Db       5604 LELVPCRAKEG------------------LEKTIADDNEPL-PDCEPTQSRHKVEEI--- 5641

Qy        771 GHIALLAPDQF---AAPWKSWVATFIVKDLLMND-RLPGKKTTKLWVPDEEVSPETMVKI 826
                    |:|       ||       |        ||:               |   :  :   |   ||                                    :|        ::
Db       5642 -DAAILRSQQFEQAADAELSWITETQKKLMSLGDIRLE-------------QDQTSAQL 5686

Qy        827 QAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDM----------- 875
              |   |          :|    |:     :    ||    :::||        |:   :|           |   |
Db       5687 QVQKAFTMDILRHKDIIDELVTSGHKIMTT----SGEEEKQSMKKKLDKVLKKYDAVCQI 5742

Qy        876 ---SRLRLAAGSAIVKLAQEPCYHEI---ITLEQ---YQLCALAINDECY--------QV 918
                 |:|       ::|       |    |   |  |:         :|    |  |  |  |:   :|                   |:
Db       5743 NSERHLQLERAQSLVSQFWE-TYEELWPWLTETQRIISQLPAPALEYETFERQQEEHRQL 5801

Qy        919 RQVFAQ------KLHK-GLSRLRL-PLE--------YMAICALCA--KDPVKERRAHARQC 961
              |:: |:          |::| |   |  |  |  |               |:|      |  :   |: ||:|
Db       5802 RELIAEHKPHIDKMNKTGPQLLELSPKEGIYIQEKYVAADTLYSQIKEDVKKRAV----V 5857

Qy        962 LVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQ---------DIE 1012
              |  :  |:      ::   :       :  |  |    :                 |  :|     :   :  :  :|  |              |:|
Db       5858 LDEAISQSTQFHDKIDQILESLERIAERLRQPPSI---SAEVEKIKEQIGENKSVSVDME 5914

Qy       1013 QLKDVKECLWFVLEILMAKNEN--------------NSHAFIRKMVENIKQTKDAQGPDD 1058
              :|:   :  |   |        |  ::|::|                               :    ||     :    :   :  ::|:  |
Db       5915 KLQPLYETLRQRGEEMIARSEGTEKDVSARAVQDKLDQMVFIWGSIHTLVEEREAKLLDV 5974

Qy       1059 AKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTK---NYLPPE 1115
               :: ||  :    ||      :: .|  |       :       :||  :                  |    :             :    |   |
Db       5975 MELAEKFW--CDHMSLVVTIKDTQDFIRDLEDPGIDIPSVVKQQQEAAEAIREEIDGLQEE 6032

Qy       1116 MKSFFT-----------PGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNP 1164
                :       |                 |  ||         :     :|      |     |   :   :   |:|           |:
Db       6033 LDMVITLGSELIAACGEPDKPIVKKSIDELNSAWDSLNKAWKDRVDRLEEAMQAAVQYQD 6092

Qy       1165 SSPG----------------------RIKGRLDSSEMDHSENEDYTM-------SSPLP 1194
                   |                                :|  :::   :    ||       |           |        :   |
Db       6093 GLQGIFDWVDIAGDKLATMSPIGTDLETVKQQIEELKQFKSEAYQQQMEMERLNHQAELL 6152

Qy       1195 GKKSDKRDDSDLVRS---ELEKPRGRKKTPVTEQEEKL------------GMDDL----- 1234
                ||   :     |       |:          ||:                :   ||                               :|:|
Db       6153 LKKVTEEADKHTVQDPLMELKLIWDSLDERIVSRQHKLEGTLLALGQFQHALDELLAWLT 6212

Qy       1235 -TK-LVQEQKPKGS--------------------------QRSRKRGHTASESDE----- 1261
                 || |:  |||| |                                    :     |  |:      ||  |
Db       6213 HTKGLLSEQKPVGGDPKAIEIELAKHHVLQNDVLAHQSTVEAVNKAGNDLIESSEGEEAS 6272

Qy       1262 ----------QQWPEEKRLKEDILENEDEQNSP------PKKGKRGR------------- 1292
                                  |:|        :||||    |::                          ||     |
Db       6273 NLQYKLRILNQRW-------QDILEKTDQRKQQLDSALRQAKGFHGEIEDLQQWLTDTER 6325

Qy       1293 ---PPKPLGG--GTPKEE 1305
                        |||||     |  ||:
Db       6326 HLLASKPLGGLPETAKEQ 6343
```

RESULT 14
A3D1_HUMAN
ID   A3D1_HUMAN   STANDARD;   PRT; 1153 AA.
AC   O14617; O00202; O75262; Q96G11; Q9H3C6;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Adapter-related protein complex 3 delta 1 subunit (Delta-adaptin 3)
DE   (AP-3 complex delta subunit) (Delta-adaptin).
GN   AP3D1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 4).
RC   TISSUE=Fetal brain;
RX   MEDLINE=97447555; PubMed=9303295;
RA   Ooi C.E., Moreira J.E., Dell'Angelica E.C., Poy G., Wassarman D.A.,
RA   Bonifacino J.S.;
RT   "Altered expression of a novel adaptin leads to defective pigment
RT   granule biogenesis in the Drosophila eye color mutant garnet.";
RL   EMBO J. 16:4508-4518(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2), AND TISSUE SPECIFICITY.
RC   TISSUE=Heart;
RX   MEDLINE=97296491; PubMed=9151686;
RA   Simpson F., Peden A.A., Christopoulou L., Robinson M.S.;
RT   "Characterization of the adaptor-related protein complex, AP-3.";
RL   J. Cell Biol. 137:835-845(1997).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).

```
RC   TISSUE=Muscle;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 33-1153 FROM N.A. (ISOFORMS 1 AND 2).
RA   Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA   Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
RA   Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA   Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA   Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA   Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA   Kronmiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA   Kobayashi A., Olsen A.S., Carrano A.V.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 74-1153 FROM N.A. (ISOFORM 1).
RC   TISSUE=Fetal liver;
RA   Zhang C., Yu Y., Zhang S., Wei H., Bi J., Zhou G., Dong C., Zai Y.,
RA   Xu W., Gao F., Liu M., He F.;
RT   "Functional prediction of the coding sequences of 75 new genes deduced
RT   by analysis of cDNA clones from human fetal liver.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PART OF THE AP-3 COMPLEX, AN ADAPTOR-RELATED COMPLEX
CC       WHICH IS NOT CLATHRIN-ASSOCIATED. THE COMPLEX IS ASSOCIATED WITH
CC       THE GOLGI REGION AS WELL AS MORE PERIPHERAL STRUCTURES. IT
CC       FACILITATES THE BUDDING OF VESICLES FROM THE GOLGI MEMBRANE AND
CC       MAY BE DIRECTLY INVOLVED IN TRAFFICKING TO LYSOSOMES.
CC   -!- SUBUNIT: ASSEMBLY PROTEIN COMPLEX 3 (AP-3) IS A HETEROTETRAMER
CC       COMPOSED OF TWO LARGE CHAINS (DELTA AND BETA3), A MEDIUM CHAIN
CC       (MU3) AND A SMALL CHAIN (SIGMA3).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
CC       Name=1;
CC         IsoId=O14617-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=O14617-2; Sequence=VSP_000165, VSP_000166;
CC       Name=3;
CC         IsoId=O14617-3; Sequence=VSP_000167;
CC       Name=4;
CC         IsoId=O14617-4; Sequence=VSP_000168;
CC   -!- TISSUE SPECIFICITY: PRESENT IN ALL ADULT TISSUES EXAMINED WITH THE
CC       HIGHEST LEVELS IN SKELETAL MUSCLE, HEART, PANCREAS AND TESTIS.
CC   -!- SIMILARITY: BELONGS TO THE ADAPTOR COMPLEXES LARGE SUBUNITS
CC       FAMILY.
CC   -!- CAUTION: Ref.3 (AAH10065) sequence differs from that shown due to
CC       the lack of 8 exons and truncation of 2 other exons in the C-
CC       terminus. Alternative splicing seems doubtful, since exon-intron
CC       junctions are not the consensus ones.
CC   -!- CAUTION: Ref.5 C-terminal sequence differs from that shown due to
CC       the presence of an unrelated sequence found on chromosome 7.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF002163; AAC51761.1; -.
DR   EMBL; U91930; AAD03777.1; -.
DR   EMBL; BC010065; AAH10065.1; -.
DR   EMBL; AC005545; AAC34212.1; -.
DR   EMBL; AC005545; AAC34214.1; -.
DR   EMBL; AF130042; AAG35473.1; ALT_INIT.
DR   Genew; HGNC:568; AP3D1.
DR   MIM; 607246; -.
DR   GO; GO:0005794; C:Golgi apparatus; TAS.
DR   GO; GO:0005480; F:vesicle transport; TAS.
DR   GO; GO:0006726; P:eye pigment biosynthesis; TAS.
DR   GO; GO:0006886; P:intracellular protein transport; TAS.
DR   GO; GO:0006899; P:non-selective vesicle transport; TAS.
DR   InterPro; IPR002553; Adaptin_N.
DR   Pfam; PF01602; Adaptin_N; 1.
KW   Golgi stack; Protein transport; Transport; Alternative splicing;
KW   Polymorphism.
FT   VARSPLIC    168    258       Missing (in isoform 2).
FT                                /FTId=VSP_000165.
FT   VARSPLIC    866    866       K -> KKAEDLDFWLSTTPPPAPAPAPAPVPSTDECEDAKT
FT                                EAQGEEDDAEGQDQD (in isoform 2).
FT                                /FTId=VSP_000166.
FT   VARSPLIC    117    285       Missing (in isoform 3).
FT                                /FTId=VSP_000167.
FT   VARSPLIC    746    877       Missing (in isoform 4).
FT                                /FTId=VSP_000168.
FT   VARIANT    1072   1072       I -> V (in dbSNP:3189383).
FT                                /FTId=VAR_014276.
FT   CONFLICT      3      3       L -> F (IN REF. 2).
FT   CONFLICT    167    167       MISSING (IN REF. 4; AAC34214).
FT   CONFLICT    252    252       E -> G (IN REF. 3; AAH10065).
FT   CONFLICT    595    603       EVSALFAGE -> DFVHCCYEL (IN REF. 5).
FT   CONFLICT    604   1153       MISSING (IN REF. 5).
FT   CONFLICT    743   1153       MISSING (IN REF. 3; AAH10065).
SQ   SEQUENCE   1153 AA;  130157 MW;  1B491DD64EAD5096 CRC64;

  Query Match             2.9%;  Score 208;  DB 1;  Length 1153;
  Best Local Similarity   18.7%;  Pred. No. 0.024;
  Matches  190;  Conservative  168;  Mismatches  412;  Indels  248;  Gaps   43;

Qy        510 LDLIKQPKTDASVKAIFSKVMVITRNLPD-----PGKAQDFMKKFTQVLEDDEKIRKQLE 564
               | ::|     |:  :||           ||  |        :|  |  :|  ::: ::  ||
Db          3 LKMVK-----GSIDRMFDK------NLQDLVRGIRNHKEDEAKYISQCIDEIKQELKQDN 51

Qy        565 VLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMI------------------KFLLE 606
               : |      ||                     :|:|:                  ||  :
Db         52 IAVKANAVCKLT-------------------YLQMLGYDISWAAFNIIEVMSASKFTFK 91

Qy        607 RIA------PVHIDTESISALIKQVNKSIDGTADDEDEG---------VPTDQAIRAGLE 651
              ||         |  |: |     |: |  :   :    | |           |  | |     :
Db         92 RIGYLAASQSFHEGTDVIMLTTNQIRKDL-SSPSQYDTGVALTGLSCFVTPDLARDLAND 150

Qy        652 LLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALL 711
              :: ::| | |      |       :|     |:     |  |:     |    |:|:  | ::|| :
Db        151 IMTLMSHTKPYIRKKA-----VLIMYKV-FLKYPESLRPAFPRLKEKLEDPDPGVQSAAV 204

Qy        712 PVLHHKSKKGPPRQAKYAIHCIHAIFSSKET----QFAQIFEPLHKSLDPSNLEHLITPL 767
              |:    ::: |           : ||         :  ::|   |    |:|   :  || ||
Db        205 NVICELARRNPKNYLSLAPLFFKLMTSSTNNWVLIKIIKLFGAL-TPLEPRLGKKLIEPL 263

Qy        768 VTIGH----IALLAPDQFAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETM 823
               : |     ::||             | | |    : ::  :|                 |
Db        264 TNLIHSTSAMSLL---------YECVNTVIAVLISLSSGMPNH-----------SASIQ 302

Qy        824 VKIQAIKMMVR---------WLLGMK---NNHSKSGTSTLRLLTTILHSDGDLTE----- 866
               : :| :::::            || |        | ||   |     |:   |      :
Db        303 LCVQKLRILIEDSDQNLKYLGLLAMSKILKTHPKSVQSHKDLILQCLDDKDESIRLRALD 362
```

```
Qy        867 --QGKISKPD--------MSRLRLAAGSA--------IVKLAQEPCYHEIITLEQY---- 904
                 | :|| :          |: :  | |:          |: :  :  |  |   | |
Db        363 LLYGMVSKKNLMEIVKKLMTHVDKAEGTTYRDELLTKIIDICSQSNYQYITNFEWYISIL 422

Qy        905 ------------QLCALAINDECYQVRQV--FAQKLHKGLSRLRLPLEYMAICALCAKDP 950
                          | |  : |   :|: :  ||      :|::   |:  : |
Db        423 VELTRLEGTRHGHLIAAQMLDVAIRVKAIRKFA------VSQMSALLDSAHLLA------ 470

Qy        951 VKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQD 1010
                 :|    : |      :  |: :          | ::|    |      |: |     | ||:
Db        471 SSTQRNGICEVLYAAAWICGEFSEHLQEPHHTLEAMLRPRVTTLPGHIQA-----VYVQN 525

Qy       1011 IEQLKDVKECLWFVLEILMAKNE----NNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLY 1066
               : :|              || |  :       : |  : ||: : |   :     | :: |:
Db        526 VVKL----------YASILQQKEQAGEAEGAQAVTQLMVDRLPQFVQSA---DLEVQERAS 573

Qy       1067 TVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPK 1126
               :  :  :|          |::   ||         :     |:    | : |: :        |:
Db        574 CILQLVKHI-------QKLQAKDVPVA-----EEVSALFAGELNPVAPKAQK--KVPVPE 619

Qy       1127 TTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNP------SSPGRIKGRLDSSEMD 1180
               ::    :|:|||     |:::   |    |           |     :   :   | :: : :
Db        620 GLDLDAWINEPLS----DSESEDERPRAVFHEEEQRRPKHRPSEADEEELARRREARKQE 675

Qy       1181 HSENEDYTMSSPLPGKKSDKR---DDSDLVRSELEKPRGRKKTPVTEQ-----EEKLGMD 1232
               : |   | ||| | |:           :       :|: :|  |        |:::|       ||:
Db        676 QANNPFYIKSSPSPQKRYQDTPGVEHIPVVQIDLSVPLKVPGLPMSDQYVKLEEERRHRQ 735

Qy       1233 DLTKLVQEQKPKGSQRSRKRGHTA--SESDEQQWPEEK--RLKEDILEN---EDEQNSPP 1285
               | |  : :| |  ::  || |::  :||||   | ::   : |:: ||     || :  |
Db        736 KLEKDKRRKKRKEKEKKGKRRHSSLPTESDEDIAPAQQVDIVTEEMPENALPSDEDDKDP 795

Qy       1286 KKGKRGRP---PKPLGGGTP---KEEPTMKTSKKGSK------KKSGPPAPEEE--EEEE 1331
                  |        |||          ::     :|||   |          |||  |  :|:  :|:|
Db        796 NDPYRALDIDLDKPLADSEKLPIQKHRNTETSKSPEKDVPMVEKKSKKPKKKEKKHKEKE 855

Qy       1332 RQSGNTEQKSKSKQHRVSRRAQQRAESPESSAIESTQSTP---QKGRGRPSKTPSPSQ 1386
               |       ::| | |  :  ::  :: :    : ::  |    ::  | |  :  :| :
Db        856 RDKEKKKEKEKKKSPKPKKKKHRKEKEERTKGKKKSKKQPPGSEEAAGEPVQNGAPEE 913

RESULT 15
NP14_HUMAN
ID   NP14_HUMAN     STANDARD;      PRT;   699 AA.
AC   Q14978; Q15030;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Nucleolar phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa
DE   nucleolar phosphoprotein) (Nopp140) (Nucleolar and coiled-body
DE   phosphoprotein 1).
GN   NOLC1 OR KIAA0035.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM ALPHA).
RC   TISSUE=Leukemia;
RX   MEDLINE=95386590; PubMed=7657714;
RA   Pai C.-Y., Chen H.-K., Sheu H.-L., Yeh N.-H.;
RT   "Cell-cycle-dependent alterations of a highly phosphorylated nucleolar
RT   protein p130 are associated with nucleologenesis.";
RL   J. Cell Sci. 108:1911-1920(1995).
RN   [2]
RP   SEQUENCE OF 3-699 FROM N.A. (ISOFORM BETA).
RC   TISSUE=Bone marrow;
RX   MEDLINE=96051387; PubMed=7584026;
RA   Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayasi Y.,
RA   Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
RT   "Prediction of the coding sequences of unidentified human genes. I.
RT   The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
RT   analysis of randomly sampled cDNA clones from human immature myeloid
RT   cell line KG-1.";
RL   DNA Res. 1:27-35(1994).
RN   [3]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=96205319; PubMed=8630004;
RA   Pai C.-Y., Yeh N.-H.;
RT   "Cell proliferation-dependent expression of two isoforms of the
RT   nucleolar phosphoprotein p130.";
RL   Biochem. Biophys. Res. Commun. 221:581-587(1996).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE=97168979; PubMed=9016786;
RA   Chen H.-K., Yeh N.-H.;
RT   "The nucleolar phosphoprotein P130 is a GTPase/ATPase with intrinsic
RT   property to form large complexes triggered by F- and Mg2+.";
RL   Biochem. Biophys. Res. Commun. 230:370-375(1997).
RN   [5]
RP   CHARACTERIZATION.
RX   MEDLINE=20036810; PubMed=10567578;
RA   Chen H.-K., Pai C.-Y., Huang J.-Y., Yeh N.-H.;
RT   "Human Nopp140, which interacts with RNA polymerase I: implications
RT   for rRNA gene transcription and nucleolar structural organization.";
RL   Mol. Cell. Biol. 19:8536-8546(1999).
CC   -!- FUNCTION: RELATED TO NUCLEOLOGENESIS, MAY PLAY A ROLE IN THE
CC       MAINTENANCE OF THE FUNDAMENTAL STRUCTURE OF THE FIBRILLAR CENTER
CC       AND DENSE FIBRILLAR COMPONENT IN THE NUCLEOLUS. IT HAS INTRINSIC
CC       GTPASE AND ATPASE ACTIVITIES. MAY PLAY AN IMPORTANT ROLE IN
CC       TRANSCRIPTION CATALYZED BY RNA POLYMERASE I.
CC   -!- SUBUNIT: INTERACTS WITH RNA POLYMERASE I 194 KDA SUBUNIT (RPA194)
CC       AND WITH CASEIN KINASE-II.
CC   -!- SUBCELLULAR LOCATION: SHUTTLES BETWEEN THE NUCLEOLUS AND THE
CC       CYTOPLASM. AT TELOPHASE IT BEGINS TO ASSEMBLE INTO GRANULAR-LIKE
CC       PRE-NUCLEOLAR BODIES WHICH ARE SUBSEQUENTLY RELOCATED TO NUCLEOLI
CC       AT THE EARLY G1-PHASE.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Alpha;
CC         IsoId=Q14978-1; Sequence=Displayed;
CC       Name=Beta;
CC         IsoId=Q14978-2; Sequence=VSP_004338;
CC   -!- PTM: THIS PROTEIN UNDERGOES RAPID AND MASSIVE PHOSPHORYLATION AND
CC       DEPHOSPHORYLATION ON CK2 AND PKC SITES. THERE IS EVIDENCE
CC       SUGGESTING THAT CDC2 KINASE PHOSPHORYLATES P130 AT THE M-PHASE.
CC   -!- SIMILARITY: Contains 1 LisH domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z34289; CAA84063.1; -.
DR   EMBL; D21262; BAA04803.1; -.
DR   PIR; I38073; I38073.
DR   Genew; HGNC:15608; NOLC1.
DR   GK; Q14978; -.
DR   MIM; 602394; -.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0005730; C:nucleolus; TAS.
DR   GO; GO:0007049; P:cell cycle; TAS.
DR   GO; GO:0007067; P:mitosis; TAS.
DR   GO; GO:0006364; P:rRNA processing; TAS.
DR   InterPro; IPR006594; LisH.
DR   Pfam; PF05022; SRP40_C; 1.
DR   SMART; SM00667; LisH; 1.
DR   PROSITE; PS50896; LISH; 1.
KW   Nuclear protein; Phosphorylation; Repeat; GTP-binding; ATP-binding;
KW   Alternative splicing.
FT   DOMAIN       10     42       LISH.
FT   DOMAIN       84    566       11 X 12 AA APPROXIMATE REPEATS OF AN
FT                                ACIDIC SERINE CLUSTER.
```

```
FT   REPEAT       84     95       ACIDIC SERINE CLUSTER 1.
FT   REPEAT      125    136       ACIDIC SERINE CLUSTER 2.
FT   REPEAT      167    178       ACIDIC SERINE CLUSTER 3.
FT   REPEAT      221    232       ACIDIC SERINE CLUSTER 4.
FT   REPEAT      264    275       ACIDIC SERINE CLUSTER 5.
FT   REPEAT      325    336       ACIDIC SERINE CLUSTER 6.
FT   REPEAT      363    375       ACIDIC SERINE CLUSTER 7.
FT   REPEAT      425    436       ACIDIC SERINE CLUSTER 8.
FT   REPEAT      470    481       ACIDIC SERINE CLUSTER 9.
FT   REPEAT      519    529       ACIDIC SERINE CLUSTER 10.
FT   REPEAT      555    566       ACIDIC SERINE CLUSTER 11.
FT   DOMAIN       68     82       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      204    382       INTERACTS WITH RPA194.
FT   DOMAIN      384    587       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      601    617       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   MOD_RES     563    563       PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT   VARSPLIC    241    241       K -> KVWTITSVRAE (in isoform Beta).
FT                                /FTId=VSP_004338.
FT   CONFLICT      3      3       D -> A (IN REF. 2).
FT   CONFLICT    133    133       R -> S (IN REF. 2).
FT   CONFLICT    291    292       YA -> SV (IN REF. 2).
FT   CONFLICT    456    456       S -> P (IN REF. 2).
SQ   SEQUENCE   699 AA;  73720 MW;  DFD4AD94EDF659FB CRC64;

  Query Match             2.9%;  Score 206.5;  DB 1;  Length 699;
  Best Local Similarity   22.3%;  Pred. No. 0.014;
  Matches   93;  Conservative   55;  Mismatches  172;  Indels   97;  Gaps   13;

Qy       1045 ENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKN 1104
              :: :: | |  |      :::      |       :: :: | :|  |      :   :| ||
Db        228 DDSEEEKAAATPKKTVPKKQVVAKAPVKAATTPTRKSSSSEDSSSDEEEEQK---KPMKN 284

Qy       1105 FSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSN- 1163
                  :| ||        |  |      ||     |  :  ||    :||  |   |: | ||:
Db        285 KPGPYSYAPP-------PSAPPPKKSLG-TQPPKKAVEKQQPVESS--EDSSDESDSSSE 334

Qy       1164 -----------------PSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKRDDSD 1205
                               | :     :   |||: | || :|   | |     |:      :
Db        335 EEKKPPTKAVVSKATTKPPPAKKAAESSSDSSDSDSSE-DDEAPSKPAGTTKNSSNKPAV 393

Qy       1206 LVRSELEKPRGRKKTPV--------------TEQEEKLGMDDLT-KLVQEQKPKGSQRS 1249
                :|    ||     | ||              :|:|      :: |  |:|    ||| : ::
Db        394 TTKSPAVKPAAAPKQPVGGGQKLLTRKADSSSSEEESSSSEEEKTKKMVATTKPKATAKA 453

Qy       1250 RKRGHTASESDEQQWPEEKR---LKEDILENEDEQNSPPKKGKRGRPPKPLGGGTP---- 1302
                    |      :| |:   |      |    :|:|:     |     : :|  |   ||  |
Db        454 ------ALSLPAKQAPQGSRDSSSDSDSSSSEEEEEKTSKSAVKKKPQKVAGGAAPSKPA 507

Qy       1303 -----------------KEEPTMKTSKKGSKKKSGPPA---------------PEEEE 1328
                               ||    |    ||| :    | |                   ||||
Db        508 SAKKGKAESSNSSSSDDSSEEEEEKLKGKGSPRPQAPKANGTSALTAQNGKAAKNSEEEE 567

Qy       1329 EEERQSGNTEQKSKSKQHRVSRRAQQRAESPESSAIE-STQSTPQKGRGRPSKTPSP 1384
              ||::::        || |  :  |       | :  ||:|::  |:   | :|   |  :     :  ||
Db        568 EEKKKAAVVVSKSGSLKKRKQNEAAKEAETPQAKKIKLQTPNTFPKRKKGEKRASSP 624

Search completed: September 24, 2003, 20:17:48
Job time : 32 secs
```

NPA

GenCore version 5.1.6

OM protein - protein search, using sw model

```
Run on:          September 24, 2003, 20:01:29 ; Search time 31 Seconds
                                          (without alignments)
                                          4315.182 Million cell updates/sec

Title:           US-09-512-581B-2
Perfect score:   7193
Sequence:        1 MAHSKTRTNDGKITYPPGVK..........QKGRGRPSKTPSPSQPKKNV 1391

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters:      283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_76:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2698.5	37.5	851	2	T00374	hypothetical prote
2	807	11.2	1579	2	T23142	hypothetical prote
3	632.5	8.8	1506	2	S52957	bimD protein - Eme
4	571.5	7.9	1303	2	E96805	hypothetical prote
5	557	7.7	1277	2	S54451	hypothetical prote
6	309.5	4.3	390	2	T43647	hypothetical prote
7	303	4.2	780	2	F96840	hypothetical prote
8	264	3.7	990	2	H86293	protein T24D18.4 [
9	235.5	3.3	1875	2	S38173	myosin-like protei
10	234.5	3.3	852	2	T06310	hypothetical prote
11	231.5	3.2	2253	2	T30336	nuclear/mitotic ap
12	229	3.2	2663	1	S28261	centromere protein
13	227	3.2	1916	2	S46157	RIF1 protein - yea
14	223.5	3.1	4717	2	T41581	hypothetical coile
15	223	3.1	2139	2	T18296	myosin heavy chain
16	222.5	3.1	1819	2	A71928	cag island protein
17	222.5	3.1	2301	2	T02323	nodulin-like prote
18	221.5	3.1	1927	2	G64585	cag pathogenicity
19	220.5	3.1	1727	2	T50073	myosin-like coiled
20	217.5	3.0	1020	1	QFHUH	neurofilament trip
21	216.5	3.0	1790	2	S67593	transport protein
22	210	2.9	1072	1	A37221	neurofilament trip
23	209.5	2.9	1365	2	T45031	hypothetical prote
24	209	2.9	1242	2	G88480	protein C16A3.7 [i
25	207.5	2.9	854	2	S02003	neurofilament trip
26	207.5	2.9	3488	2	T34418	hypothetical prote
27	206.5	2.9	699	2	I38073	nucleolar phosphop
28	206	2.9	6642	2	T29757	protein UNC-89 - C
29	205	2.8	2094	2	S33124	tpr protein - huma
30	203.5	2.8	2364	2	A56577	microtubule-associ
31	202	2.8	2401	2	T28676	rhoptry protein -
32	202	2.8	2748	2	S57976	nuclear migration
33	202	2.8	6713	2	B89921	hypothetical prote
34	201.5	2.8	3225	2	I52300	giantin - human
35	200.5	2.8	2269	2	T28677	rhoptry protein -
36	200	2.8	1837	2	T41023	probable nuclear p
37	199	2.8	1359	2	T34036	hypothetical prote
38	199	2.8	2464	1	QRMSP1	microtubule-associ
39	199	2.8	3187	2	JC5837	364K Golgi complex
40	198.5	2.8	1695	2	T19823	hypothetical prote
41	197.5	2.7	1538	2	T29095	cardiac muscle fac
42	197.5	2.7	3259	1	A56539	giantin - human
43	195.5	2.7	1679	2	S48385	hypothetical prote
44	195	2.7	488	2	I46014	cylicin II - bovin
45	194	2.7	374	2	C88734	protein F32E10.6 [

ALIGNMENTS

```
RESULT 1
T00374
hypothetical protein KIAA0648 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C;Accession: T00374
R;Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura,
DNA Res. 5, 169-176, 1998
A;Title: Prediction of the coding sequences of unidentified human genes. X. The compl
A;Reference number: Z14142; MUID:98403880; PMID:9734811
A;Accession: T00374
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-851 &lt;ISH&gt;
A;Cross-references: EMBL:AB014548; NID:g3327109; PIDN:BAA31623.1; PID:g3327110
A;Experimental source: brain
C;Genetics:
A;Note: KIAA0648

  Query Match             37.5%;  Score 2698.5;  DB 2;  Length 851;
  Best Local Similarity   62.1%;  Pred. No. 1.1e-121;
  Matches  540;  Conservative  114;  Mismatches  167;  Indels   49;  Gaps   11;

Qy        477 LYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITRNL 536
              ||||||:|| ||||||||||||||:||  |::|||| |||  ::|:  |:| |:| | :||
Db          1 LYYLYASLDPNAVKALNEMWKCQNMLRSHVRELLDLHKQPTSEANCSAMFGKLMTIAKNL 60

Qy        537 PDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNP 596
              |||||||||:||| ||| ||||:| |||:|:|||||||||: ||||| :|| ||||||||
Db         61 PDPGKAQDFVKKFNQVLGDDEKLRSQLELLISPTCSCKQADICVREIARKLANPKQPTNP 120

Qy        597 FLEMIKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVL 656
              ||||:|||||||||||||:|:||||:| :||||:||||||:|||   | |||:||||||||
Db        121 FLEMVKFLLERIAPVHIDSEAISALVKLMNKSIEGTADDEEEGVSPDTAIRSGLELLKVL 180

Qy        657 SFTHPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHH 716
              ||||| ||||||||:|||| ||:|:|:||||||:|||:||| ||| | | ||| |:|:||
Db        181 SFTHPTSFHSAETYESLLQCLRMEDDKVAEAAIQIFRNTGHKIETDLPQIRSTLIPILHQ 240

Qy        717 KSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALL 776
              |:|:| | ||| |:||||||||::|| | ||||||| :||:     | ||||||::|||::|
Db        241 KAKRGTPHQAKQAVHCIHAIFTNKEVQLAQIFEPLSRSLNADVPEQLITPLVSLGHISML 300

Qy        777 APDQFAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWL 836
              ||||||:| || || ||||||||||||   |:|  ||| |||||||| : |:||||::||||
Db        301 APDQFASPMKSVVANFIVKDLLMNDRSTGEKNGKLWSPDEEVSPEVLAKVQAIKLLVRWL 360

Qy        837 LGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYH 896
              |||||| |||  ||||||: :| |:|||||| :||| ||||||||||||||:|||||||||
Db        361 LGMKNNQSKSANSTLRLLSAMLVSEGDLTEQKRISKSDMSRLRLAAGSAIMKLAQEPCYH 420

Qy        897 EIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRA 956
```

```
Db        421 EIITPEQFQLCALVINDECYQVRQIFAQKLHKALVKLLLPLEYMAIFALCAKDPVKERRA 480

Qy        957 HARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKD 1016
Db        481 HARQCLLKNISIRREYIKQNPMATEKLLSLLPEYVVPYMIHLLAHDPDFTRSQDVDQLRD 540

Qy       1017 VKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNII 1076
Db        541 IKECLWFMLEVLMTKNENNSHAFMKKMAENIKLTRDAQSPDESKTNEKLYTVCDVALCVI 600

Qy       1077 MSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNK 1136
Db        601 NSKSALCNADSPKDPVLPMKFFTQPEKDFCNDKSYISEETRVLLLTGKPKPAGVLGAVNK 660

Qy       1137 PLSSAGKQSQTKSSRMETVS--NASSSSNPSSPGRIKGR-LDSSEMDHSENED--YTMSS 1191
Db        661 PLSATGRKPYVRSTGTETGSNINVNSELNPSTGNRSREQSSEAAETGVSENEENPVRIIS 720

Qy       1192 PLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRK 1251
Db        721 VTPVKNID--------------------PVKNKE--INSDQAT-----QGNISSDRGKK 752

Qy       1252 RGHTASESDE-QQWPEEKRLKEDILENEDEQNSP-PKKGKRGRPPKPLG-GGTPKEEPTM 1308
Db        753 RTVTAAGAENIQQKTDEK---------VDESGPPAPSKPRRGRRPKSESQGNATKNDDLN 803

Qy       1309 KTSKKGSKKKSGPPAPEEEEEEEERQSGNTE 1338
Db        804 KPINKGRKR----AAVGQESPGGLEAGNAK 829
```

RESULT 2
T23142
hypothetical protein H38K22.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T23142; T24139
R;Barlow, K.
submitted to the EMBL Data Library, June 1998
A;Reference number: Z19695
A;Accession: T23142
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1579 <WIL>
A;Cross-references: EMBL:AL024499; PIDN:CAA19710.1; GSPDB:GN00021; CESP:H38K22.1
A;Experimental source: clone H38K22
R;Ainscough, R.
submitted to the EMBL Data Library, August 1995
A;Reference number: Z19843
A;Accession: T24139
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1579 <WI2>
A;Cross-references: EMBL:Z50874; PIDN:CAA90774.1; GSPDB:GN00021; CESP:H38K22.1
A;Experimental source: clone R10E4
C;Genetics:
A;Gene: CESP:H38K22.1
A;Map position: 3
A;Introns: 28/3; 99/3; 130/3; 239/3; 323/3; 363/2; 396/1; 444/3; 559/2; 651/3; 689/3; 79

```
  Query Match            11.2%;  Score 807;  DB 2;  Length 1579;
  Best Local Similarity  21.5%;  Pred. No. 8.4e-31;
  Matches  352;  Conservative  282;  Mismatches  674;  Indels  326;  Gaps   57;

Qy         11 GKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEE---KELYLNLALHLASDF 67
Db          3 GDVDYPQGCLPILNSNNVTQQIERLRKLFNCLHDCKTNNSEEVSSPNRFARLFQHLSQEC 62

Qy         68 FLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDT--KSPQF 125
Db         63 FLDNSNTDFRILLSLCLANILRIFQPELPTPSVMDLKEVYIYLFRTMRGLGDVTQDSPKF 122

Qy        126 NRYFYLLENIAWVKSYNICFELEDSNE-----IFTQLYRTLFSV-INNGHNQ-------- 171
Db        123 KNYFSLVETME--KIIPPIIEMKDHDDKEATPVFRALIKDILAIPCGKGWNQNLKKEARL 180

Qy        172 ---------------------KVHMHMVDLMSSIICEGDTVSQELLDTVLVNLVPAHKN 209
Db        181 LKIQENDDDSMNDDEEDENAAEKIRKSLIQIATTVITNLDFVQNECLDVLFYHIINPQRS 240

Qy        210 LNKQAYDLAKALLK--------RTAQAIEPYITTFFNQVLMLGK-------TSISDLSEH 254
Db        241 NFAEARALAEDIIRSCSDNESDTLANSIRSTMTAAAKE----GKLPEEFELTGSSNRSK- 295

Qy        255 VFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMFGAKDSELA-SQNKP 313
Db        296 FFEVLRYLHYVSFDLVSGAIQELKFWLQSENEQYRKEAVTVVGMLTRDKHCQFGMDSNDP 355

Qy        314 LWQCYLGRFNDIHVPIRLECVKFASHCLM-NHPDLAKDLTEYLKVRSHDPEEAIRHDVIV 372
Db        356 TWSAFLNASIDQDDSVRHEFVQQSKDILISNHSHLRGQIINSLLRLSVDLNDDIRRDVVT 415

Qy        373 SIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYALQ---------- 422
Db        416 GVTEVAKTKLEVISDKMLKACAERMKDKKPKVRIQAIKRLMDLYNHVMTSSPQPFFSKDG 475

Qy        423 SAAGKDA----------AKQIAWIKDKLLHIY--------YQNSIDDRLLVERIFAQYM 463
Db        476 SCAPKDSKEATLSYTESEKESVRFISTAVFNVYRLTQRLPAYH---DARAIIERYFQLYL 532

Qy        464 VPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQ--PKTDAS 521
Db        533 VPYKVEPKLRVRLMADLFRNLDDIGCMLFGDIINRSSQLRRAMIGILSQVGQFTEMTTQS 592

Qy        522 VKAIFSKVMVITRNLPDPGKAQDFMKKF-TQVLEDDEK---IRKQLEVLVSPTCSCKQAE 577
Db        593 SAQLKERIRRICQIFPDAQILEKNMMVFVNQMAENDETFNLVKKLMSESYTSEENARTAG 652

Qy        578 GCVREITKKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQVNKSIDGTADDED 637
Db        653 ALQAIMEKKISSKAQQT-----VFRHFIDRIVPLSFDVPTAKEIIHLVSDTVCAKVDLKK 707

Qy        638 EGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDEKVA-EAAL----QIF 692
Db        708 W---AENCFEKDLSLLKIFTDNFGYLFADEQIIEEIRSKILASEEPIAIEAALHVLSKIF 764

Qy        693 KNTGSKIEEDFPHIRSALLPVLHHKSK--------------KGPP---RQAKYAIHCIHA 735
Db        765 ANS---------HFRNKLENEATHKEKWFLLLGKDLKDLVMREEPELRRSCKLATRLLSF 815

Qy        736 IFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPWKSWVATFI-- 793
Db        816 ILGKEKV--IEFFD--------DQIEQLISRLYIESQGAANAFQVLGEIFRCDISYYLPQ 865

Qy        794 VKDLLMNDRL-PGKKTTKLWVPDEEVS-----------PE-TMVKIQAIKMMVRWLLGM 839
Db        866 VMDVVESEKIGPMILTSSMHGNDDPVEFNELMHIEKQPWPKYAMAKVYAAKFATKVLTVY 925

Qy        840 --------KNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQ 891
Db        926 PLIPSTEDKRRMEKAAQNFIDLLSEIIEKKGDL--GGKQCDLEQARLRATASGCLLKLAS 983

Qy        892 EPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAI-----CALC 946
Db        984 VITYRTKLNTHIFKNMSYIITDEAYCVRLYYALHVKKGLSKNRLPIEFAACYGLVNLGLS 1043

Qy        947 AKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSL--------LPEYVVPYTIHL 998
Db       1044 EEDGENKMDGFKTICM----NQAQQAFGERNDNQATLLKLEGPQRAIFCSESVIAYVVWL 1099

Qy        999 LAH-----------------DPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIR 1041
Db       1100 LANYDKLEKVEGNANRNDSSEELEIKVANVNLLSELQESLWLVIDSLKIAKCNMQKVW-- 1157

Qy       1042 KMVENIKQTKDAQGPDDAKM-----------------NEKLYTVCDVAMNIIM------- 1077
```

```
Db    1158 KVLEKLKTCGDKSMRSDSSLSTTRLREHNEVNPDSRGNIKLWVLCDLGITMMLYRAKLQM 1217

Qy    1078 ----SKSTTYSLE-----SPKDPVLPARFFTQPDKNFSNTKN-----------YLPPEMK 1117
Db    1218 EDQEAKEAGFNLQFFYVCSPKDKADPSNVYA-PDVLINDEKNRNGRIPKPGRVFHVSDLT 1276

Qy    1118 SFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSS 1177
Db    1277 SEFTP--PPQGN---------ESTGSNSSKNASRRANVTGTGRKRGGAGTKTTKRKSGGS 1325

Qy    1178 EMDHSENEDYTMSSPLPGKKSDKRDDSDLVRSELEK------PRGRKKTPVTEQEEKLGM 1231
Db    1326 KISDDEGDSMDVKSPAVKKTRSKRGEYDLPDEEDEEMEVIPLPKRRGAAPDSTIASSSSN 1385

Qy    1232 DDLTKLVQEQKP---------KGSQRSRKRGHTASESDEQQWPEEKRLKEDILENEDEQN 1282
Db    1386 GSILKNGSGPSPKKRNSRGSVKGRQNSTKEAEIDVDSDEEM--------EEITEKRDNVS 1437

Qy    1283 ------SP---PKKGKRGRPPKPLGGGT--PKEEPTMKTSKKGSKKKSGPPAPEEEEEEE 1331
Db    1438 LDNLIISPILDESSGRTRRSARTIAATATITSSTPLVPTKPKITKRKRSEAVVEEVEDEE 1497

Qy    1332 RQSGNTEQKSKSKQHRVSRR---------AQQRAESPESSAI--ESTQSTPQKGR-GRP- 1378
Db    1498 PEI--NDKKSPSPKKRVSGRRSAPTPTKNSKKAPVSPKKKTLIKETNGVSPKKNKYGLPM 1555

Qy    1379 -------SKTPSPS 1385
Db    1556 EEDDDSIGKTPKPT 1569
```

RESULT 3
S52957
bimD protein - Emericella nidulans
C;Species: Emericella nidulans, Aspergillus nidulans
C;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 07-May-1999
C;Accession: S52957
R;Denison, S.H.; Kaefer, E.; May, G.S.
Genetics 134, 1085-1096, 1993
A;Title: Mutation in the bimD gene of Aspergillus nidulans confers a conditional mitotic
A;Reference number: S52957; MUID:93387663; PMID:8375649
A;Accession: S52957
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1506 <DEN>
A;Cross-references: EMBL:L03200; NID:g168024; PID:g168025

```
  Query Match             8.8%;  Score 632.5;  DB 2;  Length 1506;
  Best Local Similarity   19.9%;  Pred. No. 1.8e-22;
  Matches  308;  Conservative  298;  Mismatches  626;  Indels  319;  Gaps   53;

Qy      30 EMVRRLKMVVKTFMDMDQDSEEEKELYLNLALHLASDFFLKHPGKDVRLLVACCLADIFR 89
Db      47 DLLQRLQTLAQELPNVEQE-EIEKESLRKVSQELATAQLLAHKDKGVRAWATCCIVDVLR 105

Qy      90 IYAPEAPYTSPDKLKDIFMFITRQ-LKGLEDTKSPQFNRYFYLLENIAWVKSYNICFELE 148
Db     106 LCAPDAPFTA-NQLKDIFTCIVSSIIPALGDPSNPYNAQHIYVLNSLAEVKSIVLMTDLD 164

Qy     149 DSNEIFTQLYRTLFSVINNGHN--------QKVHMHMVDLMSSIICEGDTVSQELLDTVL 200
Db     165 HPDTLIVPLFISCFDIVAGSAKASTGEPVAKNVEYDMTRLLVTVIDESPVLAPDVVDVIV 224

Qy     201 VNLV---------PAHKNLNKQ-----------------AYDLAKALLKRTAQAIEPYIT 234
Db     225 AQFLRVDPRVLDGPGKKGKKPETQVDEKQETLLLKDYPAAYNMAKAICQACPERMTSHIS 284

Qy     235 TFFNQVLM--------------LGKTSISDLSEHVFD---------LILELYNIDSHLLL 271
Db     285 QYFNNVIIDASATGTQNGPSKQARRTNLDDSDEEGEDIKELSKAHRLIRELWRACPDVLQ 344

Qy     272 SVLPQLEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPL----------------- 314
Db     345 NVIPQLEAELSAESVSLRLLATQTIGDL--TSGTGVAGPPPPLPMDPAVYPQVKLDDYAR 402

Qy     315 -----------------------WQCYLGRFNDIHVPIRLECVKFASHCLMN------- 343
Db     403 SIPQPNVLLMPFAPKPFSQAHSSAYDSFLSRRLDKSASVRASWRTAIGRIILTSAGGSGL 462

Qy     344 HPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDI---------LLVNDHLLNFVR 394
Db     463 SDNEEQTLITHLSSMLRDADERVRLAAVEAVGTFGLSHIVNKLGVSGGVSTQDSLLFILA 522

Qy     395 ERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWI----KDKLLHIYYQNSID 450
Db     523 ERVKDRKSQVREHATKVLA---RAWAVASGDIERSHEQVTPLLKEAPSRILDAYYTNDPE 579

Qy     451 DRLLVERIFAQYMVP------------------HNLETTE-------------RMKCLYY 479
Db     580 IHVSIDRAMFEILLPLSYPPIKPKLSRSSSSQSQRLKDSQAAEPESEADVDRIRVRRILT 639

Qy     480 LYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQ------PKTDASVKAIFSKVM-VI 532
Db     640 LVGGLDEKAKKVFFAMQKRQVSLRTAVTVYLQACEEYNGGVMEKNKDQIKAQLTKIVDAL 699

Qy     533 TRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQ 592
Db     700 AKTFPDPARTSADLWKFAKI--HDRRGYQLIRFAMAAVSDYRTVIKAIKELARRL---QS 754

Qy     593 PTNPFL-EMIKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLE 651
Db     755 SNNTILHETLTTLLYRCSSIVFNRSHIPA-IMSISRS-----DENGLAAPAH-------E 801

Qy     652 LLKVLSFTHPISFHSAETFESLL--ACLKMD-------------DEKVAEAALQIFKNTG 696
Db     802 MLKEISSLNP------EVLEAQVQEICKDLEAQAPKATTVSAAGTEEILKACSGFAKKLP 855

Qy     697 SKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLD 756
Db     856 SKLPKERKFFQALVDYALHSPS----PRAAKHAVLILMAV-TDKKNMYAKDLVEKCVSKC 910

Qy     757 PSNLEHLITPLVTIGHIALLAPDQFAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDE 816
Db     911 TYDSERFLTKLATLSQLNLLAPRE-ADEESDAIIKISVNQILLTNRSPTPNSGYFW--SD 967

Qy     817 EVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTST-----------LRLLTTILHSDGDLT 865
Db     968 QVDDETAAKEWALKIIVNRL------RAKDGSDSDDDFRAHAEPVYDTLNKLIVNSGELS 1021

Qy     866 EQGKISKPDMSRLRLAAGSAIVKL-AQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQ 924
Db    1022 KKKDTPATQKSRLRLLAANSLLKLCSSHALCEQLLTPQDFNSIALVAQDPLPEVRSGFIN 1081

Qy     925 KLHKGLSR-LRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHA------ 977
Db    1082 QLKKKLVQDTRLGARWYVIPYLLAFEP----QVGLKDSTLTWLRSRAAFFSQQTNGKKGE 1137

Qy     978 --AVSEKLLSLLPEYVVPYTIHLLAHDPDY--VKVQDIEQLKDVKECLWFVLEILMAKNE 1033
Db    1138 KQTVMEALFSRL--------LSLLAYHPDYPPADLDESTKLDELTDFARYILFYLSAVAN 1189

Qy    1034 NNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVL 1093
Db    1190 EHNLSLIFHIAQRVKQARDGITKSD-EMSRRLHTLSDLAQATIRRFADVYSQQ------- 1241

Qy    1094 PARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSS--- 1150
Db    1242 --RRFGGGAGGVNLLQTY----------PGK---VGVPSSIFAPMGSHREAQEVADTTFL 1286

Qy    1151 ------RMETVSNAS-SSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKR-D 1202
Db    1287 PEDAEDRLDRIVRATMRTKNGSSQAGAKKRKTDSTQEPSRDANAAKKARKNSESSSRRKS 1346

Qy    1203 DSDLVRSELEKPRGRKKT--PVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESD 1260
```

```
Db       1347 GSSAVGFKMPKRKSTKKTGDDWSSDGEAAGNVASSSATRRRSNRGS-ASRRISYADPDSD 1405

Qy       1261 E--QQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKK 1318
              |   :  |  : ::|  | ||:               | :  |:   |      |:  |
Db       1406 EDDMEMDELNQARDDEDEGEDQ------------AKDIENGSDLSE----LSEADSNML 1448

Qy       1319 SGP-----PAPEEEEEEERQSGNTEQKSKS-----KQHRVSRRAQQRAESP 1359
                |      |: :|:: :::|:|: : :  |        : :|   :| ::|  |
Db       1449 EEPEDDDGPSEKEDQPDDKQNGDDDAQPASPVPVASKAKVPGKAMKKATLP 1499
```

RESULT 4
E96805
hypothetical protein T5M16.19 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: E96805
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: E96805
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1303 <STO>
A;Cross-references: GB:AE005173; NID:g6382504; PIDN:AAF07790.1; GSPDB:GN00141
C;Genetics:
A;Gene: T5M16.19
A;Map position: 1

```
  Query Match             7.9%;  Score 571.5;  DB 2;  Length 1303;
  Best Local Similarity   20.6%;  Pred. No. 1.2e-19;
  Matches  294;  Conservative  270;  Mismatches  566;  Indels  297;  Gaps   53;

Qy         38 VVKTFMDMDQDSEEEKELYLN--------LALHLASDFFLKHPGKDVRLLVACCLADIFR 89
                |   |  :|| |   || |             |    :      ||:   || |||  |::::||
Db         21 VANTLSKIDQPSATNKEKGLKLIEAELRPLKKSIIKHALLKNRDNDVSLLVTVCVSELFR 80

Qy         90 IYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSYNICFEL-- 147
              | ||  |:   :  : |:||     :  | || || |::    :|| :: :|      |  :
Db         81 ILAPHLPF-EDEYLRDIFTLFIAEFSELSDTVSPYFSKRAKILETVSRLK---FCLLMLD 136

Qy        148 EDSNEIFTQLYRTLFSVINNGH-----NQK----------------VHMHMVDLMSSII 185
               ||  ::  :::   ||::    |     |||                 :  ::: :|| ::
Db        137 EDCQDLVHEMFNMFFSLVREHHQQSLINQKSMKTQQRKANTQQTQHSLFNNILAIMSDVL 196

Qy        186 CEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTFFNQVLMLGK 245
                |     :    :  :| |||    ::       |   || :|::|  |   :|| | :|      |
Db        197 EE--EANSSFVVVILENLVKEGEDTTSGADKLASSLIERCADRLEPLICSFLTSCFMEKD 254

Qy        246 TSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMFGAKDS 305
               :  ::| :   ::| ::  |    :||:|:|:|  :| ::   :  |:: :  |   ::|
Db        255 SIQTNLKDSYHEIIFKISLIAPQMLLAVIPKLTQELLTDQVDVRIKALNLAGRIFAQPKH 314

Qy        306 ELAS---QNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHP--DLAKDLTEYLKVRSH 360
              |:|     :  |:  :| ||:|   :|:    :|  :   |   :   :  :|     |
Db        315 CLSSYVETYQDLYAEFLRRFSDKSAEVRMAALKCGKQCYFANPSGNKASGVLTAIQERLL 374

Qy        361 DPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYA 420
              | :: :|    ::           | ::   |   :|::    ||   ||:   |||:|:   | ::|: |
Db        375 DFDDRVRTQALIVACDIMKFNMKYVPLNLISEASERLRDKKISVRKKALQKLTEVYQDYC 434

Qy        421 LQSAAG----KDAAKQIAWIKDKLLHIYYQNSIDD--RLLVERIFAQYMVPHNLETTERM 474
               : : |        |  :||       |:|  :  :  : ::       :| :  :  :  |  |   |||
Db        435 DKCSEGDMTITDNFEQIPC---KILLLCCEKNCEEFRSQNLELVLSDDLFPRLLPVEERM 491

Qy        475 KCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTD------ASVKAIFSK 528
              :     :| ::    :|:|| :    :   |:::::  | | :: | |            |: | |
Db        492 RHWVQCFAIMNHIHLKSLNSILSQKRRLQNELRHCLTLWRKAKVDNIEEAQRKKKSYFVK 551

Qy        529 VMVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLG 588
              ::      ||    :|:|  :|  ::     |   |   |  :|:      |    |:     ::  |  |:
Db        552 ---LSACFPDASEAEDLFEKLDRM--RDASIFDVLTLLLE-ELSSTNAQIIKEKFLKMIG 605

Qy        589 NPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRA 648
                    :    |  ::  |   : :|      :|  :   |: |:   |                  |
Db        606 ----VKHSLFEFLRILSTKCSPSIFSSEHVQCLLNQLCGSTSANT---------QLKAP 651

Qy        649 GLELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRS 708
               ::|| |:    |         :|    :    |   |: :::   |:   :  :          |   :|
Db        652 SIKLLLVILNMFPSYLRGSE--KQFLKLLE-ENDSAADELIVVLSKAAPYISVNF----G 704

Qy        709 ALLPVLHHKSKKGPPRQAKYAIHCIHAIF-SSKETQFAQIFEPLHKS-LDPSNLEHLITP 766
                |||       :|    |  |  |:  |  ::   ||:::  |:::   |  |   |  |     |:   :
Db        705 DYYPVLEKVCLEGTRSQTKCAVSAISSLAGSSEKSVFSELCEMLMDSLLCGRNIPTTLQS 764

Qy        767 LVTIGHIALLAPDQFAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKI 826
                |  :|   ::|  |                    |  :|:          |:  ::              :  :|
Db        765 LACVGQYSVLEYDN-----------IYEDI---------TSYIY---------RVFQI 793

Qy        827 QAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGD-------LTEQGKISKPDMSRLR 879
               :|  :|:  |         :        |  :|   |  |  |             |      |:   :  :|
Db        794 YGLKTLVKSFLPRHGQVVRKIDDLLNILKKTLKSQGHDGIKSWCLFVLEICSEDTGANVR 853

Qy        880 LAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLH--KGLSRLRLPL 937
               |||  |:: |:::    |                      |:  |  :::    :  |:       |
Db        854 LAAAKAVLLLSRKWDLH-------------ISPEVFRLTILMAKSFRYINGF------- 892

Qy        938 EYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIH 997
                           :  :      :|  |                       |  |   ::::               |  |:   :  :|
Db        893 ------------INKATRESRTC---------RDLDQGESLTDS-----PVYMTVFLIH 925

Qy        998 LLAHDPDYVKVQDIEQLKD----VKEC--LWFVLEILMAKNEN-----NSHAFIRKMVEN 1046
              :|||||::        |   :|     :  |    |: ||::|:: |  |          |:  :
Db        926 VLAHDPEFPS----EDCRDEHIYARFCGPLFSVLQVLLSINNNGFTIKETAPFLFCIFRA 981

Qy       1047 IKQTKDA----QGPDDAK---MNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFT 1099
              ||: :||      :  |:  ||      |      :      :   ::|    |: : :          || :
Db        982 IKRAEDAVDSRKTPNKAKSRTRNALEQSFIERIVHIFQSQISMHDQRCQKDSLA------ 1035

Qy       1100 QPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMET----- 1154
                      :      |||  :       |       |:: |            |:     |:  |:  ||  |
Db       1036 -----VGSEDKVLPPLL------GNQIETSITG------STEASQNNTRCSRKRTHLGEH 1078

Qy       1155 ----------------------------------VSNASSSSNPSSPGRIKGRLDSSEMD 1180
                                                |||  :||  |   |   |  :      |  |
Db       1079 ISCNSLSLRTVESEIPIKKLERHTTCAKESVKASVSNKITSSKHS--GVVSALKDIS--N 1134

Qy       1181 HSE---NEDYTMSSPLPG-------KKSDKRDDSDLV---RSELEKPRGRKKTPVTEQEE 1227
               |   |      :  |:  ||    |            :|  :  :  :  :|  :           ::          |    |
Db       1135 HGEAIIGQRIKLLSPTDGCFYPGTVEKFNSKSNSHKIIFDNGDVELVCLDSESWETLSHE 1194

Qy       1228 KLGMDDLTKLVQEQKPKGSQRS-RKRGHTASESDEQQWPEEKRLKEDILENEDEQNSPPK 1286
               :|    :   :|  :|  :   ||:        :  || ::      |:   :        |:     :  | |
Db       1195 SMGQQE--RLGKETESYGSRNCVPEISHTLAKVTAQK--QTTTKQQNKKVPAKLNPPAA 1250

Qy       1287 KGKRGRPPKPLGGGTPKE--------EPTMKTSKK------GSKKKS 1319
               | |:|      | |:  |             :     :  ||         |:||||
Db       1251 KSKKGNSDS--GEGSVSEVTDTSDNIKSECRLIKKSIAQFSGTKKKS 1295
```

RESULT 5
S54451
hypothetical protein YMR076c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YM9582.01c; hypothetical protein YM9916.15c
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jul-1995 #sequence_revision 09-Mar-1996 #text_change 29-Oct-1999
C;Accession: S54451; S52836
R;Gentles, S.; Bowman, S.

```
submitted to the EMBL Data Library, May 1995
A;Reference number: S54451
A;Accession: S54451
A;Molecule type: DNA
A;Residues: 1-682 <GEN>
A;Cross-references: EMBL:Z49259; NID:g807956; PIDN:CAA89222.1; PID:g807957; MIPS:YMR076c
A;Experimental source: strain AB972
R;Pearson, D.; Bowman, S.
submitted to the EMBL Data Library, April 1995
A;Reference number: S52814
A;Accession: S52836
A;Molecule type: DNA
A;Residues: 659-1277 <PEA>
A;Cross-references: EMBL:Z48952; NID:g763008; PIDN:CAA88801.1; PID:g763023; MIPS:YMR076c
A;Experimental source: strain AB972
C;Genetics:
A;Gene: SGD:PDS5
A;Cross-references: SGD:S0004681; MIPS:YMR076c
A;Map position: 13R

  Query Match              7.7%;  Score 557;  DB 2;  Length 1277;
  Best Local Similarity   20.3%;  Pred. No. 5.8e-19;
  Matches  272;  Conservative  267;  Mismatches  611;  Indels  190;  Gaps   46;

Qy         12 KITYPPGVKEISDK-ISKEEMVRRLKMVVKTFMDMDQDSEEEKEL--YLNLALHLASDFF 68
               |: :   :    ||: ||  |:: |||  :      :|||: :    |  |:     ||
Db          8 KLKFNSPIISTSDQLISTNELLDRLKALHEELASLDQDNTDLTGLDKYRDA---LVSRKL 64

Qy         69 LKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRY 128
              |||     :|   ||||:||  |:|||:||||    :|  |||  :   |  |  | ::    :
Db         65 LKHKDVGIRAFTACCLSDILRLYAPDAPYTDA-QLTDIFKLVLSQFEQLGDQENGYHIQQ 123

Qy        129 FYLLENIAWVKSYNICFELEDSNEIFTQLYR-----------TLFSVINNGHNQKVHMHM 177
               ||:  :    :|  :  :|  || :  :|:                ||:||
Db        124 TYLITKLLEYRSIVLLADLPSSNNLLIELFHIFYDPNKSFPARLFNVIGG---------- 173

Qy        178 VDLMSSIICEGDTVSQELLDTVL-------VNLVPAHKNLNKQ-AYDLAKALLKRTAQAI 229
               ::   :|  |  |:|  |:|    :             |    |:::  |   :   :
Db        174 --ILGEVISEFDSVPLEVLRLIFNKFLTYNPNEIPEGLNVTSDCGYEVSLILCDTYSNRM 231

Qy        230 EPYITTFFNQVLML-----GKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSN 284
               ::|  ::::::          :  :  :  :  |:|  |:        |:  :|:  :  :|  |
Db        232 SRHLTKYYSEIIHEATNDDNNSRLLTVVVKLHKLVLRLWETVPELINAVIGFIYHELSSE 291

Qy        285 DNEERLQVVKLLAKMFGA-KDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMN 343
               :   |  :  ||: ::  :    |     |   :  :: : :  ||    :|:|   :     :
Db        292 NELFRKEATKLIGQILTSYSDLNFVSTHSDTFKAWISKIADISPDVRVEWTESIPQIIAT 351

Qy        344 HPDLAKDLTEYLKVRSHDPEEAIRHDVIVSI----VTAAKKDIL--LVNDHLLNFVRERT 397
               |::|:|  :  |        |  :   :|    ::        ||    |:|      :    ||:  ||
Db        352 REDISKELNQALAKTFIDSDPRVRRTSVMIFNKVPVTEIWKNITNKAIYTSLLHLARE-- 409

Qy        398 LDKRWRVRKEAMMGLAQIYKKYALQ---SAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLL 454
                |    ||:  :   :|:  |        :    :    |:  :  |  |    |  :::||  |  ::
Db        410 --KHKEVRELCINTMAKFYSNSLNEIERTYQNKEIWEIIDTIPSTLYNLYYINDLNINEQ 467

Qy        455 VERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIK 514
               |: :  :|::|    :    :|:   |  :    |   |            |    :     :   :|  |
Db        468 VDSVIFEYLLPFEPDNDKRVHRLLTVLSHFDKKAFTSFFAFNARQIKISFAISKYIDFSK 527

Qy        515 QPKTDASVKAIFSKVMV---------ITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEV 565
                    |: :    :::            :   |  |    ||  |  ::      |      :||:|    |
Db        528 FLNNQESMSSSQGPIVMNKYNQTLQWLASGLSDSTKAIDALETIKQF--NDERIFYLLNA 585

Qy        566 LVSPTCSCKQAEGCVREITKKLGNPK-------------QPTNPFLEMIKFLLERIAPVH 612
               |:            :  |   |:  ||   |                     |   :    ::|:  ||  |  |  :|:
Db        586 CVTNDIPFLTFKNCYNELVSKLQTPGLFKKYNISTGASIMPRD-IAKVIQILLFRASPII 644

Qy        613 IDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLE--LLKVLSFTHPISF-HSAET 669
                :  :||  |:    ||                    |   :    |:    :|   :|    :|    |
Db        645 YNVSNISVLLNLSNNS--------------DAKQLDLKRRILDDISKVNPTLFKDQIRT 689

Qy        670 FESLLACL------KMDDEKVAEAALQIFKNTGS-KIEEDFPHIRSALLPVLHHKSKKGP 722
               :::: |       | |:  :  ||    ::|  :  :  |  :  ||       :      |:  :  :
Db        690 LKTIIKDLDDPDAEKNDNLSLEEALKTLYKASKTLKDQVDFDD--TFFFTKLYDFAVESK 747

Qy        723 PRQAKYAIHCIHAIFSSKETQFAQI---FEPLHKSLDPSNLEHLITPLVTIGHIALLAPD 779
                |   |||   | |:     |      :|      ||      |         |:|   :            :   |
Db        748 PEITKYATKLI-ALSPKAEETLKKIKIRILPLDLQKDKYFTSHIIVLMEIFKKFPHVLND 806

Qy        780 QFAAPWKSWVATFIVKDLLMNDRLPGKKTTKL-WVPDEEVSPETMV----KIQAIKMMVR 834
                 :  :  ::::|::|:::::  |        :    ||  |    :|            |:   :|:
Db        807 D-----STDIISYLIKEVLLSNQVVGDSKKEIDWVEDSLLSDTKYSAIGNKVFTLKLFTN 861

Qy        835 WLLGM-----KNNHSKSGT-STLRLLTTILHSDGDL-TEQGKISKPDMS----RLRLAAG 883
                |    :     ::  :  :  ::|   |  |::|  :|  |:|  :|  |  |   |   |    |    :||  ||
Db        862 KLRSIAPDVPRDELAESFTEKTMKLFFYLIASGGELISEFNKEFYPTPSNYQTKLRCVAG 921

Qy        884 SAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAIC 943
                ::|||:    :   |                |  ||     ||:  |  ::|     ::    :  ::::  :
Db        922 IQVLKLARISNLNNFIKPSDIIKLINLVEDESLPVRKTFLEQLKDYVANELISIKFLPLV 981

Qy        944 ALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDP 1003
                 |  :|   |  :    :    :   |         :     |:              :  |   :|   ||  :||  |
Db        982 FFTAYEPDVELKTTTKIWI--NFTFGLKSFKK---------GTIFERALPRLIHAIAHHP 1030

Qy       1004 DYVKVQDIE------QLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIK------QTK 1051
               |  |   |  |          |       :  |  |   :  :  |:     :      ::  :  |:|   :           :
Db       1031 DIVGGLDSEGDAYLNALTTAIDYLLFYFDSIAAQENFSLLYYLSERVKNYQDKLVEDEID 1090

Qy       1052 DAQGPDDAK-------MNEKLYTVCDVAMNIIMS----KSTTYS-----LESPKDPVLPA 1095
               : :||    :                  :|:|  :  ::|   |:::          |  :  :|           |    |    |
Db       1091 EEEGPQKEEAPKKHRPYGQKMYIIGELSQMILLNLKEKKNWQHSAYPGKLNLPSDLFKP- 1149

Qy       1096 RFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETV 1155
               |  |   :     |  |  |:|    :                 :          |:     :  :  |  ::    |  |          |:
Db       1150 -FATVQEAQLS-FKTYIPESLT------EKIQNNIKAKIGRILHTSQTQRQRLQKRLLAH 1201

Qy       1156 SNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSD-KRDDSDLVRSELEKP 1214
               |    |                   :  :  |   |  |        |         |    |   |::  |:    ::|    ::|
Db       1202 ENNESQKKKKKVHHARSQADDEEGDGDRESDSDDDSYSPSNKNETKKGHENIV---MKKL 1258

Qy       1215 RGRKKTPVTEQEEKLGMDDL 1234
               |  ||:     :  |:     ||:
Db       1259 RVRKEVDYKDDED----DDI 1274
```

```
RESULT 6
T43647
hypothetical protein - fission yeast (Schizosaccharomyces pombe) (fragment)
C;Species: Schizosaccharomyces pombe
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C;Accession: T43647
R;Lee, M.; Yoo, H.S.; Chung, K.S.
submitted to the EMBL Data Library, February 1998
A;Description: Clone 17 (bimD homologue).
A;Reference number: Z22596
A;Accession: T43647
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-390 <LEE>
A;Cross-references: EMBL:AF049529; PIDN:AAD02493.1

  Query Match              4.3%;  Score 309.5;  DB 2;  Length 390;
  Best Local Similarity   24.8%;  Pred. No. 8.4e-08;
  Matches  105;  Conservative   76;  Mismatches  192;  Indels   51;  Gaps   11;

Qy        825 KIQAIKMMVRWLLGMKNNHS--KSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAA 882
               |: ||:::|   |                 |     ::||   :|  :|||:|:           |      :  |||  |
Db          4 KVLAIRVLVNRLRAAAGGTEALNIGAPIIKLLKVLLMADGELSPFKNTPKISRAYLRLTA 63

Qy        883 GSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAI 942
               :||    |  |  |  |          |     :  ||  :  ||  :|  |  |  |  |    :||:  |      :
Db         64 SKYFLKLCSIPFYAEHIDFSSYVSNLLIMPDENFDVRNLFLTKLQKQLQLKKLPISYYPL 123

Qy        943 CALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHD 1002
```

```
                 | | || :| :  |            :::   |  :|      |||  | ||||:|
Db        124 LFLTAVDPEEEIKTKA----------SIWIRSQVAFFQKTHDFTMEYVATYLIHLLSHH 172

Qy       1003 PDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMN 1062
              ||   ::     |  :   ::|  :: :|       :  :::  |||: |    :|   |
Db        173 PDISSIESENSLDFIAYIRFYVDTVVNSENV----PIVFHLMQRIKQSYDV--IEDG--N 224

Qy       1063 EKLYTVCDVAMNIIMSKS-------TTY--SLESPKDPVLPARFFTQPDKNFSNTKNYLP 1113
               :| : |:|  |:  ||       |||   :: | : : |     :  : |:  | ::
Db        225 NYIYVLSDMAQKILQVKSQNFGWSLTTYPKQIKLPYEILRPIPSIDEKKRIFN--KIFIT 282

Qy       1114 PEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGR 1173
              |:|:|            :  |:  |:||  ||:  | : ::      || |:  |  | |
Db        283 PKMES---------QIEHAIRTPVSSFAKQTTNKHANLKQKKTHSSKSDKKSSRRRK-- 330

Qy       1174 LDSSEMDHSENEDYTMSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDD 1233
                 :|      ||        :| : | :      :  ||          ::|:||:: :|
Db        331 ---NEKRRKLNEQNPNIRNVPERSSSRFQGIRINYSEAPS----SSEEISEEEEEISEED 383

Qy       1234 LTKL 1237
               ::
Db        384 FDEI 387
```

RESULT 7
F96840
hypothetical protein F23A5.16 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: F96840
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: F96840
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-780 <STO>
A;Cross-references: GB:AE005173; NID:g6503292; PIDN:AAF14668.1; GSPDB:GN00141
C;Genetics:
A;Gene: F23A5.16
A;Map position: 1

```
  Query Match              4.2%;  Score 303;  DB 2;  Length 780;
  Best Local Similarity   20.4%;  Pred. No. 4.3e-07;
  Matches  153;  Conservative  118;  Mismatches  254;  Indels  224;  Gaps   27;

Qy         18 GVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLAL-----HLASDFFLKHP 72
              |   :|   | ::::  |              :::||   : |  :  ||       | |   | ||
Db         18 GTNLLSPPSSTDDLLTLLDETESLLKNVEQD----QPLSMQSALIPSRNALVSVDLLSHP 73

Qy         73 GKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRYFYLL 132
               |||: |  || :| || ||| ||  | |  | :|:||         :  | |   |   :  : ::|
Db         74 DSDVRVSVVSCLTEIVRITAPETPY-SDDLMKEIFRLTIEAFEKLADASSRSYKKAEFVL 132

Qy        133 ENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSSIICEGDTVS 192
              :|:| |||  :  :||    ::   |::|  |   |  :  |  |  |     |  :| :|| | : ||
Db        133 DNVAKVKSCLVMLDLE-CYDLILQMFRNFFKFIRSDHPQLVFSSMELIMIAIIDETEQVS 191

Qy        193 QELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTFFNQVLMLGKTSISDLS 252
               :|||::|  :   ::|::  :: ||: :|  |  |: ::|||      : |      ||:   |
Db        192 TDLLDSLLATVKKENQNVSPMSWSLAEKVLSRCARKLKPYII----EALKSRGTSLDMYS 247

Qy        253 EHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQV------VKLLAK-------- 298
                |   :  ::|                |::   : : ::|    ||          :: |
Db        248 PVVSSICQSVFN---------TPKVHSPVNTKEHESLKQVRSESTDAEITGKRGRKPNSL 298

Qy        299 ---------MFGAKDSELASQNKPLWQ------CYLGRFNDIHVPIRLECVKFASHCLM 342
                         : | :|    | || : :          ||:        |:  |
Db        299 MNPEDYDISWLSGKRDPLKTSSNKKIQKKGSGGVSSLGKVPAKKTPLPKE---------- 348

Qy        343 NHP-DLAKDLTEYLK--------------------------VRSHDPEE---------- 364
              | |    :: || ||                              |  :| :
Db        349 NSPATSSRSLTGSLKRSRVKMDESDYDSDSLSSPRLKKLASCFRDEEPNQEDDRKIGNSS 408

Qy        365 ---AIRHDVIVSIVTAAKKDIL---LVND-----HLLNFVRERTLDKRWRVRKEAMMGLA 413
                 :: :   |  || || ::   :||          :  : | |:   |   : :: :
Db        409 KQTRSKNGLEKSQKTAKKKPVVEAKIVNSSGKRLSARSVAKRRNLE---RAPLDTLVPQS 465

Qy        414 QIYKKYALQSAA-----------------------------GKDAAKQ---IAWIKDK 439
                ||  |  ||                                 |:|    :   | |  ||
Db        466 SKRKKMVSQVAARQLANESEEETPKSHPTRRRTVRKEVESDGFGEDLVGKRVNIWWPLDK 525

Qy        440 LLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQ 499
                 :|:  ||                            | |    :|:  |    :     | |
Db        526 ---TFYEGVID-----------------SYCTRKKMHRVIYSDGDSEELNLTEERW--- 561

Qy        500 NLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLEDDEKI 559
                         :|::   :     | |        :||:     | |::          :|:
Db        562 -----------ELLEDDTSADEDKEI---------DLPESIPLSDIMQR--------QKV 593

Qy        560 RKQLEVLVS--PTCS--------CKQAEGCVREITKKLGNPKQPTNPFLEMIKFLLERIA 609
               :|    | ||   || |              : | : : |::   :: |           | :
Db        594 KKSKNVAVSVEPTSSSGVRSSSRTLMKKDCGKRLNKQVEKTREGKN---------LRSLK 644

Qy        610 PVHIDTESISALIKQVNKSIDGTADDEDE 638
               :: :|:  :     ::    |::  :|| |
Db        645 ELNAETDRTA---EEQEVSLEAESDDRSE 670
```

RESULT 8
H86293
protein T24D18.4 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: H86293
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H86293
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-990 <STO>
A;Cross-references: GB:AE005172; NID:g6587800; PIDN:AAF18491.1; GSPDB:GN00141
C;Genetics:
A;Gene: T24D18.4
A;Map position: 1

```
  Query Match              3.7%;  Score 264;  DB 2;  Length 990;
  Best Local Similarity   17.5%;  Pred. No. 4.3e-05;
  Matches  241;  Conservative  188;  Mismatches  440;  Indels  506;  Gaps   50;

Qy         43 MDMDQDSEEEKELYLNLALHLASDFFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDK 102
              :: |  |  :| |:   :    :::|  |::|  |||: |  || :| || ||||||  : ::
Db         47 VEQDLSSSVQKALHPPMRALVSAD-LLRNPDSDVRVSVVSCLTEIMRITAPEAPY-NDEQ 104

Qy        103 LKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLF 162
              :||||       : | |   |  : :   :|| :| |:| |:|     :  :||    ::: ::::
Db        105 MKDIFQVTIEAFEKLADASSRSYRKAEVILETVAKVRSSLVMLDLE-CDDLVLEMFQRFL 163

Qy        163 SVINNGHNQKVHMHMVDLMSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALL 222
```

```
Db     164 KIIRPDHPQLVLVSMETIMITVIDESEEVPMDLLEILLTTVKKDSQDVSPAALTLVEKVL 223

Qy     223 KRTAQAIEPYITTFFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEF--- 279
Db     224 SSCTCKLQPCIM----EALKSSGTS-----------LDMYS----PVVSSICQSEFATT 263

Qy     280 ------KLKSNDNEERL---QVV--KLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHVP 328
Db     264 QAHNDVKPKDNEADEKISEGQVVPSDSLEDKLNLGLSRKGTRSKRSARGGTRRANGD--- 320

Qy     329 IRLECVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDH 388
Db     321 ---EKVITANEGLSESTD-AETASGSTRKRGWKPKSLMNPEEGYSFKTSSSKKV------ 370

Qy     389 LLNFVRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNS 448
Db     371 ----------------QEKELG---------DSSLGKVAAKKVP------------- 389

Qy     449 IDDRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKD 508
Db     390 --------------LPSKVGQTNQ------------SVVISLSSSGRARTGSRKRSR- 420

Qy     509 LLDLIKQPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLEDDEK--IRKQLEVL 566
Db     421 ----TKMEETDHDVSSVATQ---------PAKKQT-VKKTNPAKEDLTKSNVKKHEDGI 465

Qy     567 VSPTCSCKQA--EGCVREITKKLGNPKQPTNPFLE--MIKFLLERIAPVHIDTESISALI 622
Db     466 KTGKSSKKEKADNGLAKTSAKK---------PLAETMMVKPSGKKL--VHSDAK------ 508

Qy     623 KQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDE 682
Db     509 ---KKNSEGASMD----TPIPQSSKS---------------------------KKKDS 532

Qy     683 KVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKET 742
Db     533 RATTPATK----------------KSEQAPKSHPKMKR----------------IAGEEV 560

Qy     743 QFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPWKSWVATFIVKDLLMNDR 802
Db     561 E-------------SNTNEL------------------------------------GEE 570

Qy     803 LPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDG 862
Db     571 LVGKRVNVWWPLDKKFYEGVIKSYCRVKKM----------HQ-----------VTYSDG 608

Qy     863 DLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVF 922
Db     609 DV-EELNLKK---ERFKIIEDKSSASEDKEDDLLESTPLSAF------------------ 646

Qy     923 AQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEK 982
Db     647 ----------------------------------------------IQREKSKKRKIVSKN 661

Qy     983 LLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRK 1042
Db     662 --------VEP------SSSPEVRSSMQTMKKKD------------------------S 682

Qy    1043 MVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPD 1102
Db     683 VTDSIKQTKRTKGALKAVSNEP-------------ESTTGKNLKSLK------KLNGEPD 723

Qy    1103 KNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSS 1162
Db     724 KTRGRT--------------GKKQ----------------KVTQAMHRKIEK-------- 745

Qy    1163 NPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPV 1222
Db     746 -----------DCDEQEDLETKDEEDSLKL-GKESDA--EPDRMEDHQELPENHNVETK 790

Qy    1223 TEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESDEQQWPEEKRLKEDILE--NEDE 1280
Db     791 TDGEE-----------QEAAKEPTAESKTNGEEPNAEPETDGKEHKSLKEPNAEPKSDGE 839

Qy    1281 QNSPPKK--------GKRGRPPKPLGGGTPKEEPTMKTS----KKGSKKKSGPPAPEEEE 1328
Db     840 EQEAAKEPNAELKTDGENQEAAKELTAERKTDEEEHKVADEVEQKSQKETNVEPEAEGEE 899

Qy    1329 EEERQSGNTEQKSKSKQHRVSRRAQQRAESPESSAIESTQSTPQKGRGRPSKTPS 1383
Db     900 QKSVEEPNAEPKTK-----VEEKESAKEQTADTKLIEKEDMSKTKGEEIDKETYS 949
```

RESULT 9
S38173
myosin-like protein MLP1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YKR095w; protein YKR415
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 21-Jul-2000
C;Accession: S38173; S40647; S31207
R;Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantale
submitted to the Protein Sequence Database, March 1994
A;Reference number: S38158
A;Accession: S38173
A;Molecule type: DNA
A;Residues: 1-1875 <BAL>
A;Cross-references: EMBL:Z28320; NID:g486586; PID:g486587; MIPS:YKR095w
A;Experimental source: strain S288C
R;Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.;
Yeast 9, 1349-1354, 1993
A;Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chr
A;Reference number: S40644; MUID:94205265; PMID:8154186
A;Accession: S40647
A;Molecule type: DNA
A;Residues: 1-1875 <BOU>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A;Title: A new yeast gene with a myosin-like heptad repeat structure.
A;Reference number: S31207; MUID:93247549; PMID:8483450
A;Accession: S31207
A;Molecule type: DNA
A;Residues: 1-300,'A',302-1875 <KOE>
A;Cross-references: EMBL:L01992; NID:g171958; PIDN:AAA34783.1; PID:g171959
C;Genetics:
A;Gene: SGD:MLP1
A;Cross-references: SGD:S0001803; MIPS:YKR095w
A;Map position: 11R

```
  Query Match             3.3%;  Score 235.5;  DB 2;  Length 1875;
  Best Local Similarity   19.5%;  Pred. No. 0.0023;
  Matches  301;  Conservative  246;  Mismatches  602;  Indels  391;  Gaps   72;

Qy      19 VKEISDKISKE------EMVRRLKMVVKTFMDMDQDSEEEKEL-YLNLALHLASDFFLKH 71
Db     278 IKGLSDSLNSEKQEFSAEMSLKQRLVDLLESQLNAVKEELNSIRELNTAKVIADDSKKQT 337

Qy      72 PG-----KDVRLL---VACCLADIFRI-----YAPEAPYTSPDKLKDIFMFITRQLKGLE 118
Db     338 PENEDLLKELQLTKEKLAQCEKECLRLSSITDEADEDNENLSAKSSSDFIFLKKQLIKER 397

Qy     119 DTK---SPQFNRYFYLLEN-IAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVH 174
Db     398 RTKEHLQNQIETFIVELEHKVPIINSFK---ERTDMLENELNNAALLLEHTSNEKNAKVK 454

Qy     175 MHMVDLMSSIICEGD--TVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPY 232
Db     455 ELNAKNQKLVECENDLQTLTKQRLD--LCRQIQYLLITNSVSNDSKGPLRKEEIQ----- 507

Qy     233 ITTFFNQVLMLGKTSI--SDLSEHVFDLILELYNI-----DSHLLLSVLPQLEFKLKSND 285
Db     508 ---FIQNIMQEDDSTITESDSQKVVTERLVEFKNIIQLQEKNAELLKVVRNLADKLESKE 564

Qy     286 NEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHP 345
```

```
Db   565 KKSKQSLQKI--------ESETVNEAK-----------EAIITLKSEKMDLESRI----E 601

Qy   346 DLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDI-LLVND---HLLNFVRERTLDKR 401
Db   602 ELQKELEE-LKT-SVPNEDASYSNVTIKQLTETKRDLESQVQDLQTRISQITREST-ENM 658

Qy   402 WRVRKEAMMGLAQIY-KKYALQSAAGKDAAKQI-AWIKDKLLHIYYQNSIDDRLLVERIF 459
Db   659 SLLNKE----IQDLYDSKSDISIKLGKEKSSRILAEERFKLL----SNTLD--------- 701

Qy   460 AQYMVPHNLETTERMKCLYYLYATL---DLNAVKALNEMWKCQNLLRHQVKDLLDLIKQP 516
Db   702 -----LTKAENDQLRKRFDYLQNTILKQDSKTHETLNEYVSCKSKLSIVETELLNL---- 752

Qy   517 KTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCK-- 574
Db   753 KEEQKLRVHLEKNLKQELNKLSPEK--DSLRIMVTQLQTLQKEREDLLEETRKSCQKKID 810

Qy   575 QAEGCVREITKKLGNP-------KQPTNPFLEMIKFLLERIAPVH-----------IDTE 616
Db   811 ELEDALSELKKETSQKDHHIKQLEEDNNSNIEWYQNKIEALKKDYESVITSVDSKQTDIE 870

Qy   617 SISALIKQVNKSID------GTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETF 670
Db   871 KLQYKVKSLEKEIEEDKIRLHTYNVMDETI-NDDSLRKELEKSKI-NLTD--AYSQIKEY 926

Qy   671 ESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKS------------ 718
Db   927 KDLY-----------ETTSQSLQQTNSKLDESFKDFTNQIKNLTDEKTSLEDKISLLKEQ 975

Qy   719 -----------KKGPPRQA---KYAIHC-------IHAIFSSKETQFAQI-----FEPLH 752
Db   976 MFNLNNELDLQKKGMEKEKADFKKRISILQNNNKEVEAVKSEYESKLSKIQNDLDQQTIY 1035

Qy   753 KSLDPSNLEHLITPLVTIGHIALLAPDQFAAPWKSWVATF-IVKDLLMNDRLPGKKTTKL 811
Db  1036 ANTAQNNYEQELQKHADVSKTISELREQLHT-YKGQVKTLNLSRDQLENAL---KENEKS 1091

Qy   812 WVPDEEVSPE-------------TMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTIL 858
Db  1092 WSSQKESLLEQLDLSNSRIEDLSSQNKLLYDQIQIYTAADKEVNNSTNGPGLNNILITLR 1151

Qy   859 HSDGDLTEQGKISKPDMSRLRLAAGSAIVKL--AQEPCYHEIITLEQYQLCALAINDECY 916
Db  1152 RERDILDTKVTVAERDAKMLRQKISLMDVELQDARTKLDNSRVEKENHSSIIQQHDD--- 1208

Qy   917 QVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREY---- 972
Db  1209 ---------IMEKLNQLNL---------------LRESNITLRNELENNNNKKKELQSEL 1244

Qy   973 --LKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEI--- 1027
Db  1245 DKLKQNVAPIESELTAL----------------KYSMQEKEQELKLAKE------EVHRW 1282

Qy  1028 ------LMAKNENNSHAFIRKM---VENIK---QTKDAQGPD-DAKMN-------EKLYT 1067
Db  1283 KKRSQDILEKHEQLSSSDYEKLESEIENLKEELENKERQGAEAEEKFNRLRRQAQERLKT 1342

Qy  1068 VCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKN-FSNTKNYLPPEMKSFFTPGKPK 1126
Db  1343 ------SKLSQDSLTEQVNSLRDA-----------KNVLENSLSEANARIEELQNAKVAQ 1385

Qy  1127 TTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENED 1186
Db  1386 GNNQLEAIRKLQEDAEKASRELQAKLE-------ESTTSYESTINGL----------NEE 1428

Qy  1187 YTMSSPLPGKKSDKRDDSDLVRSELEKPRG-RKKTPVTEQEEKLGMDDLTKLVQEQKPKG 1245
Db  1429 IT-----------------TLKEEIEKQRQIQQQLQATSANEQ---NDLSNIVESMK--- 1465

Qy  1246 SQRSRKRGHTASESDEQQWPEEK--RLKEDILENEDEQNSPP-------KKGKRGRPPKP 1296
Db  1466 ---------KSFEEDKIKFIKEKTQEVNEKILEAQERLNQPSNINMEEIKKKWESEHEQE 1516

Qy  1297 LGGGTPKEEPTMK------TSKKGSK----KKSGPPAPEEEEEEER-----QSG------ 1335
Db  1517 VSQKIREAEEALKKRIRLPTEEKINKIIERKKEELEKEFEEKVEERIKSMEQSGEIDVVL 1576

Qy  1336 ------NTEQKSKSKQHRVSRRAQQRAES-PESSAIESTQ 1368
Db  1577 RKQLEAKVQEKQKELENEYNKKLQEELKDVPHSSHISDDE 1616
```

RESULT 10
T06310
hypothetical protein F11C18.80 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999
C;Accession: T06310
R;Bevan, M.; Terryn, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;
ewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, April 1999
A;Reference number: Z15589
A;Accession: T06310
A;Molecule type: DNA
A;Residues: 1-852 <BEV>
A;Cross-references: EMBL:AL049607; GSPDB:GN00062; ATSP:F11C18.80
A;Experimental source: cultivar Columbia; BAC clone F11C18
C;Genetics:
A;Gene: ATSP:F11C18.80
A;Map position: 4
A;Introns: 34/3; 102/3; 163/2; 205/3; 269/3; 282/3; 564/3; 602/3; 622/2; 639/3; 667/3

```
  Query Match             3.3%;  Score 234.5;  DB 2;  Length 852;
  Best Local Similarity   16.9%;  Pred. No. 0.00092;
  Matches  236;  Conservative  175;  Mismatches  391;  Indels  591;  Gaps   49;

Qy    18 GVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLN-LALHLASDFFLKHPGKDV 76
Db    16 GEKLIDPPSSLDELLSFLDKLFVSLAEVEQSPPDSMQNALTPLMKGLVGGKLFKHSDVDV 75

Qy    77 RLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIA 136
Db    76 KVAVAACISEITRITAPDAPY-DDDQMKEVFKLIVSSFEDLVDKSSRSYAKRISILETVA 134

Qy   137 WVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSSIICEGDTVSQELL 196
Db   135 KVRSCVVMLDLE-CDALLIEMFQHFLKAIRDHHSGNVFSSMENIMTLVLEESEDIPSEML 193

Qy   197 DTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYIT-TFFNQVLMLGKTS--ISDLSE 253
Db   194 SPIL-HSVKKDDEISQVSRRLAEQVLSNCASKLKTYLTEAVKSSGVPLDKYSNIVASICE 252

Qy   254 HVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKP 313
Db   253 GTF---------------SALQQ--DQVVANEKED--------------------SQG-- 273

Qy   314 LWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVS 373
Db   274 ------------HIKRETEVEKAAEISTPERTDAPKDESGKSGVSN---GVAQQNDSSVD 318

Qy   374 IVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQI 433
Db   319 TDSMKKQD-----------------------------------------DTGAKDEPQQL 337

Qy   434 AWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALN 493
Db   338 ----DNPRNTDLNNTTEEKPDVE---------HQIEEKENESS---SVKQADLSKDSDIK 381

Qy   494 EMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVL 553
Db   382 EETEPAELL-----DSKDVLTSPPVDSSVTAATS------------------------- 410

Qy   554 EDDEKIRKQLEVLVSPTCSCKQA---------EGCVREITKKLGNPKQPTNPFLEMIKFL 604
```

```
Db        411 SENEK-NKSVQILPSKTSGDETANVSSPSMAEELPEQSVPKKTANQKKKESS-------- 461

Qy        605 LERIAP-VHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPIS 663

Db        462 TEEVKPSASIATEEVS-------------------------------------------- 477

Qy        664 FHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPP 723

Db        478 -----------------EEPNTSEPQVTKKSGKKVA------------------------ 496

Qy        724 RQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAA 783

Db        497 -------------SSSKTK---------------------PTVP---------------- 506

Qy        784 PWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNH 843

Db        507 -------------------PSKKSTS---------ETKVAKQSEKKVVG-----SDNA 531

Qy        844 SKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQ 903

Db        532 QES---------------------TKPKEEKKKPGRGKAI-------------------- 550

Qy        904 YQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLV 963

Db        551 ---------DE---------ESLHTSSGDNEKP-------AVSSGKLASKSKKEAKQTVE 585

Qy        964 KNIN---VRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKEC 1020

Db        586 ESPNSNTKRKRSLGQGKASGESLVG------------------SRIKV------------ 615

Qy       1021 LWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKS 1080

Db        616 -WWPMD----------QAYYKGVVESY----------DAAKKKHLVIYDDGDQEILYLKN 654

Qy       1081 TTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSS 1140

Db        655 QKWS-----------------------------------------------------PLDE 662

Qy       1141 AGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDK 1200

Db        663 S-ELSQDEEAADQTGQEEDASTVGSGAGSSKAK-----------------ATPASKSSK 703

Qy       1201 --RDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPK-----GSQRSRKRG 1253

Db        704 TSQDDKTASKSKDSKEASREEEASSEEESE----------EEEPPKTVGKSGSSRSKK-- 751

Qy       1254 HTASESDEQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKK 1313

Db        752 ------------------DI-------SSVSKSGKSKASSK------KKEEPSKATT-- 777

Qy       1314 GSKKKSGP----PAPEEEEEEERQSGN-TEQKSKSKQHRVSRRAQQRAESPESSAIESTQ 1368

Db        778 SSKSKSGPVKSVPAKSKTGKGKAKSGSASTPASKAKESASESESEETPKEPE----PATK 833

Qy       1369 STPQKGRGRPSKT 1381

Db        834 AKSGKSQGSQSKS 846
```

RESULT 11
T30336
nuclear/mitotic apparatus protein - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C;Accession: T30336
R;Merdes, A.; Ramyar, K.; Vechio, J.D.; Cleveland, D.W.
Cell 87, 447-458, 1996
A;Title: A complex of NuMA and cytoplasmic dynein is essential for mitotic spindle assem
A;Reference number: Z20828; MUID:97053784; PMID:8898198
A;Accession: T30336
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-2253 <MER>
A;Cross-references: EMBL:Y07624; NID:g1514670; PIDN:CAA68905.1; PID:g1514671
C;Genetics:
A;Gene: NuMA

```
  Query Match             3.2%;  Score 231.5;  DB 2;  Length 2253;
  Best Local Similarity   19.2%;  Pred. No. 0.0046;
  Matches  329;  Conservative  248;  Mismatches  535;  Indels  601;  Gaps   80;

Qy          7 RTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSE---------------- 50

Db        272 RENQTELQEPKELEELREK--NESLMIRLRDTLKQCQDMKADKKLLERKNDQLAEENGEL 329

Qy         51 ---------------------EEKELYL--------------------------NLAL 61

Db        330 SYKVRDLSNRLAQLQEALYETTEEQELSLSNWQQKQNQLESELSGAVGEKKYLEEHNLIL 389

Qy         62 H----LASDFFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIF-------MFI 110

Db        390 QGKISMLED-QLKEMGEIDMPETGDCMGDILKL----------DDLKQELAVLNTQCLSL 438

Qy        111 TRQLKGLEDTKS-------PQFNRYFYLLENIAWVKSYNICFELEDS--NEIFTQLYRTL 161

Db        439 KEQIHQMEEEKSTAEVEMEAQKSR-----------------FESEKGQLQEIVTNLQTSL 481

Qy        162 FSVI-------NNGHNQKVHMHMVDLMSSIICEGDTVSQEL--LDTVLV----------N 202

Db        482 SEITFQKERLDNEARAQQEH---------LMCQITTLKLEISKLKSSLVHKDEELKGIHH 532

Qy        203 LVPAHKNLNKQAYDLAKAL---LKRTAQAIE------PYITTFFNQVLMLGKTSISDLSE 253

Db        533 KVEEERNEKNQLLENFKMLGGNLIGITQQLESKTKEVDYLREQQQKILCERDSTLSTLNE 592

Qy        254 H---------VFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMFG--A 302

Db        593 YKCKKDENSGVLNKTVKTLEQDHQTSLSVIE----KLKSEKEELASKVQDLDAKMLGLIA 648

Qy        303 K----DSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVR 358

Db        649 KCQNLDSENDSQSK-------------------------SHAA----------TVESLKAQ 674

Qy        359 SHDPEEAIR-------HDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMG 411

Db        675 LSEQESQLKIYRKKVSSNELVSEENSKLKDQLLSVEESLRHLREH-------------- 719

Qy        412 LAQIYKKYALQSAAGKDA-----------AKQIAWIKDKLLHIYYQNSIDDRLLVERIFA 460

Db        720 ----LEKEKTKFAASLDADLKRISHLEEEMKKLSESRDEALH----NLDEERTAGKKIES 771

Qy        461 QYMVPHNLETTERMKCLYYLYATL--DLNAVKALNE----------MWKC------QNLL 502

Db        772 QL---KHLE-EEYQKANESLQAKLAGSCAAIKQREEERDELSKVVDIWKAKYGESQQKIA 827

Qy        503 RH--QVKDLLDLIKQPKTDA--SVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLED--- 555

Db        828 QNSCHMQEQTEELKKTHSDVYQQLEGERSKVLMIEAKASETKSSQ--LEKINQLEGELSA 885

Qy        556 ---------------------DEKIR-------KQLEVLVSPTCSCKQAEGCVREITKK 586

Db        886 ANACIKEREAEEKKLVSALHSAEEKLKIAYQGESERLSHLETALSNAKQDLDC---LAKE 942

Qy        587 LGNPKQPTNPFLEMIKFL----LERIAPVH-------------IDTESISALIKQVNKS 628

Db        943 LSDEKYKKAEFEAMVKVLKEQNSERIASLESELKNSLAVVKERKCESEKLSGEVEHLKRQ 1002

Qy        629 IDGTADDEDEGV-----------------PTDQAIRA--GLELLKVLSFTHPISFHSAET 669

Db       1003 LDDSSQKHKEALAQKNIEIKQLINAKEKATSDLAIKSEMGAQLQKAVD-THKSEFSALQN 1061

Qy        670 FES----LLA-------------------------CLKMDDEKVAEAAL---------- 689

Db       1062 ELSRSLDLLALKEGEVERLNKEAALRQEEIQQQQTITKLTEEETALAALKDKVALQEKE 1121

Qy        690 ---QIFKNTGSKIEEDFPHIRSALLPVLHHKSKK-----GPPRQAKYAIHCIHAIFSSK- 740
```

```
               |:    |:  |::   ::|   |:  |||:          :  |  : ||     ||
Db      1122 IKQQVQATKGA--EKEMAKLKS----VISEKSKRIECLEQDIQNQKRDLSCIQEQHQSKL 1175

Qy       741 -ETQFAQ---------------IFEPLHK-----------SLDPSNLEHLITPLVTIG 771
              |:|  |                 | |  :|            |:     || :    : ||
Db      1176 GESQGLQALIADLEKKCKEQKELICEAQNKAAEAKTLASEKASVSERQLEGIQALEIEIG 1235

Qy       772 HIALLAPD---QFAAPW-----KSWVATFIVKDLLMN-DRLPGKKTT--------KLWVP 814
               |  |      |       |           : |:|     |   :|:          |::
Db      1236 KERQKACDLQKQLELSWAVQEEKETELQALKKELFHKVQELEQSQTSFTDSSGEALLYLS 1295

Qy       815 DEEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGD---------LT 865
             : :    : : : :        :  : |||     |   :  :::|:: : :: :          |
Db      1296 EAQERQQALTEAKEQAEQYQKEIEMKNKEVNSLQAEIKILSSKVTTNEEVSVDFEQRLLK 1355

Qy       866 EQGKISKPD--MSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFA 923
             |   | :| :   | :|  :     ::    :|  :: |   : :|  |      |     |  |:
Db      1356 ETSKSAKLEEKMQKLHMELEASFKELLEKNCAIDCLTTEAQNLKGEA--DQQRMAVDSLQ 1413

Qy       924 QKL-HKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEK 982
             |||   |   :    |   |   |      ||:    ||::  :  |    :|:     :   |:: |::
Db      1414 QKLSSKAETNHTLQQEIQAWQKNCAE---KEQQICSLQ---QNLKSNQSLLEEFASLKHS 1467

Qy       983 LLSLLPEYVVPYTIH---LLAHDPDYVKVQ-DIEQLK-DVKECLWFVLEILMAKNENNSH 1037
              :: |   :    | ||:|          :  |  ::|:  |  |:  |  :  :|:   :    |     |
Db      1468 YQEIIAERDLMQEKHQEELLSHKKLTERFQAELEKAKEDMTEIV--LLKEKLHNQELQLH 1525

Qy      1038 AFIRKMVENIKQTKDAQGPDDAKM--NEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPA 1095
              |   :   :: |       |    :    :     |:  |    :  |        :|:  :|      |   |:
Db      1526 KFQSENSYSLTQISHLQQVNSQLLGANQSLSQISDQGAKKLESEMSTLK-EQHKE----- 1579

Qy      1096 RFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETV 1155
                                  |||:                    | |      || |   | ::|||
Db      1580 ------------------EMKTLRL----------QYEKTLREGNKQVQETSLQLETV 1609

Qy      1156 SNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKRDDSDLVRSELEKPR 1215
                                   :|:  ||                           |:|:: |  :
Db      1610 -------------------TSKYDH-----------------------VKSKVLKDQ 1624

Qy      1216 GRKKTPVTEQEEK----LGMDDLTK-LVQEQKPKGSQRSR---KRGHTASESDEQQWPEE 1267
                       |  ||||     |   :  :|  |  |  |::|    ||:  :     :  |   |   |:|:
Db      1625 ------KTFQEEKQRLLLQVQELNKQLSQQEKTIRSQQQKLKQREGETHEEADK----SH 1674

Qy      1268 KRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAPEEE 1327
              ||: |  ||:: ||  :    :   :  :    |                 |           ||::        :|
Db      1675 KRVLE--LESQLEQQTQAVEHYKAQMEK-----------AKVHYDAKKKQN------QE 1714

Qy      1328 EEEERQSGNTEQKSKSKQHRVSRRAQQRAESPE 1360
               || ||    :|:   ||::      |   :||| :
Db      1715 LSEELQSHIKQQEHLSKEN-----ADLKAESEQ 1742
```

RESULT 12
S28261
centromere protein E - human
N;Alternate names: centromere 312K protein; kinesin-related protein CENP-E
C;Species: Homo sapiens (man)
C;Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C;Accession: S28261
R;Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
Nature 359, 536-539, 1992
A;Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.
A;Reference number: S28261; MUID:93024922; PMID:1406971
A;Accession: S28261
A;Molecule type: mRNA
A;Residues: 1-2663 <YEN>
A;Cross-references: EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865
C;Genetics:
A;Gene: GDB:CENPE
A;Cross-references: GDB:361164; OMIM:117143
A;Map position: 4q24-4q25
C;Superfamily: centromere protein E; kinesin motor domain homology
C;Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
F;7-335/Domain: kinesin motor domain homology <KMOT>
F;86-93/Region: nucleotide-binding motif A (P-loop)
F;486-2183/Domain: coiled coil #status predicted <COI>
F;92/Binding site: ATP (Lys) #status predicted

```
  Query Match             3.2%;  Score 229;  DB 1;  Length 2663;
  Best Local Similarity   18.2%;  Pred. No. 0.0076;
  Matches  275;  Conservative  268;  Mismatches  537;  Indels  434;  Gaps   64;

Qy        17 PGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEE---------EKELYLNLALHLASD 66
             | | |:|   : | :::| :   |  ||: :  ||             ||:     || |
Db       333 PYVNEVS---TDEALLKRYR---KEIMDLKKQLEEVSLETRAQAMEKD---QLAQLLEEK 383

Qy        67 FFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIF-MFITRQLKGL-EDTKSPQ 124
             |:                                       :|::::  | :|     | :: |: :
Db       384 DLLQK------------------------VQNEKIENLTRMLVTSSSLTLQQELKAKR 417

Qy       125 FNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSSI 184
               |  : |  |  :|: |                 |    |::  |     | |    ::|:  |
Db       418 KRRVTWCLGKINKMKNSN--------------YADQFNIPTN-ITTKTHKLSINLLREI 461

Qy       185 ---ICEGDTVSQELLDTVL-VNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTFFNQV 240
                :|      |      |||:   :    || | ||::   :
Db       462 DESVCSESDVFSNTLDTLSEIEWNPATKLLNQENIE---------------------- 497

Qy       241 LMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMF 300
                 : :: |       :|:|      |     |  :     ::| |||   :: :   : ::     |
Db       498 -----SELNSLRADYDNLVL-----DYEQLRTEKEEMELKLKEKNDLDEFEALERKTK-- 545

Qy       301 GAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSH 360
               || |:                               |::        |:| :: :|: :
Db       546 --KDQEMQ-------------------------------LIHEISNLKNLVKHREVYNQ 571

Qy       361 DPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYA 420
              | |   :           ::| ::|    :   :  ::|      ::      :    | |:         |:
Db       572 DLENEL---------SSKVELLREKEDQIKKLQEYIDSQK---LENIKMDLS-----YS 613

Qy       421 LQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYL 480
             |:|        ||   :  :  :                |   :  :  ::   |||  |:||  |
Db       614 LESIEDPKQMKQTLFDAETV-----------ALDAKRESAFLRSENLELKEKMKELATT 661

Qy       481 YATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITR------ 534
              |  ::        |::     |:  |   :  |   :||  |:        ::: |:::  :|
Db       662 YKQME-------NDIQLYQSQLEAKKKMQVDLEKE------LQSAFNEITKLTSLIDGKV 708

Qy       535 ------NLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLG 588
                      ||    ||   |     |:    :  :|::|:  ::|::: :|:|        |:  :::
Db       709 PKDLLCNLELEGKITDLQKELNKEVEENEALREEV-ILLS---ELKSLPSEVERLRKEI- 763

Qy       589 NPKQPTNPFLEMIKFLLERIAPVHIDTES-ISALIKQVNKSIDGTADDEDEGVPTDQAI- 646
                 |   :   | :|        :::       :   ||  :   |::::  |: |   |    :        |||
Db       764 ---QDKSEELHIITSEKDKLFSEVVHKESRVQGLLEEIGKTKDDLATTQSNYKSTDQEFQ 820

Qy       647 --------------------RAGLELLKVLSFTHPISFHSAETFESLLACLKMD----- 680
                                     |     |:: :                |: |:|  |  || :
Db       821 NFKTLHMDFEQKYKMVLEENERMNQEIVNLSK--------EAQKFDSSLGALKTELSYKT 872

Qy       681 ---DEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKS-----------------K 719
                 ||   |     ::   |    :::|    :   |  |   |                           :
Db       873 QELQEKTREVQERL--NEMEQLKEQLENRDSPLQTVEREKTLITEKLQQTLEEVKTLTQE 930

Qy       720 KGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSN-----LEHLITPLVTIGHIA 774
               |    :|  :  ::      ::     :  | : :: ::|              |||   |        ||  :
Db       931 KDDLKQLQESLQI------ERDQLKSDIHDTVNMNIDTQEQLRNALESLKQHQETINTL- 983

Qy       775 LLAPDQFAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVR 834
                             || ::   : ::|  |                           || : ||   : |        :
Db       984 ----------KSKISEEVSRNLHM----------------EENTGETKDEFQ------Q 1010

Qy       835 WLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISK--PDMSRLRLAAGSAIVKLAQ- 891
               ::|:                  :| : ||   :   :   : :| ||              : :  |:     | |  :  |
Db      1011 KMVGIDKKQDLEAKNT-QTLTADVKDNEIIEQQRKIFSLIQEKNELQQMLESVIAEKEQL 1069
```

```
Qy        892 -----EPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALC 946
Db       1070 KTDLKENIEMTIENQEELRL----LGDELKKQQEIVAQEKNHAIKK---EGELSRTCDRL 1122

Qy        947 A--KDPVKERRAHARQCLVKNINVRREYLKQHAAVSE--------KLLSLLPEYVVPYTI 996
Db       1123 AEVEEKLKEKSQQLQEKQQQLLNVQEEMSEMQKKINEIENLKNELKNKELTLEHMETERL 1182

Qy        997 HLLAH-DPDYVKVQDI-EQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQ 1054
Db       1183 ELAQKLNENYEEVKSITKERKVLKE-----LQKSFETERDHLRGYIREIEATGLQTKEEL 1237

Qy       1055 GPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPP 1114
Db       1238 KIAHIHLKEHQETIDELRRSV--SEKT------------AQIINTQDLEKSHTK--LQE 1280

Qy       1115 EMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKS--SRMETVSNASSSSNPSSPGRIKG 1172
Db       1281 EIP-----------VLHEEQELLPNVKKVSETQETMNELELLTEQSTTKDSTTLARI-- 1326

Qy       1173 RLDSSEMDHSE-NEDYTMSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTE------- 1224
Db       1327 -----EMERLRLNEKFQESQEEIKSLTKERDNLKTIKEALEVKHDQLKEHIRETLAKIQE 1381

Qy       1225 ----QEEKLGM----DDLTKLVQEQ---KPK-----------GSQRSRKRGH-----TA 1256
Db       1382 SQSKQEQSLNMKEKDNETTKIVSEMEQFKPKDSALLRIEIEMLGLSKRLQESHDEMKSVA 1441

Qy       1257 SESDEQQ------WPEEKRLKEDI-------LENEDEQNSPPKKGKRGRPP-KPLGGGTP 1302
Db       1442 KEKDDLQRLQEVLQSESDQLKENIKEIVAKHLETEEELKVAHCCLKEQEETINELRVNLS 1501

Qy       1303 KEEPTMKTSKK-----GSKKKSGPPAPEEEEEE---ERQSGNTEQKSKSKQHRVSRRAQQ 1354
Db       1502 EKETEISTIQKQLEAINDKLQNKIQEIYEKEEQLNIKQISEVQENVNELKQFKEHRKAKD 1561

Qy       1355 RA-ESPESSAIEST 1367
Db       1562 SALQSIESKMLELT 1575
```

```
RESULT 13
S46157
RIF1 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YBR1743; protein YBR275c; RAP1-interacting factor
C;Species: Saccharomyces cerevisiae
C;Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 21-Jul-2000
C;Accession: S46157; S46156; S44537; S25704; S39131
R;Brandt, T.; Christiansen, C.; Holmstroem, K.; Kallesoe, T.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S46157
A;Accession: S46157
A;Molecule type: DNA
A;Residues: 1-1096 <BRA>
A;Cross-references: EMBL:Z36144; MIPS:YBR275c
A;Experimental source: strain S288C
R;Aigle, M.; Baclet, M.C.; Barthe, C.; Biteau, N.; Crouzet, M.; Doignon, F.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45940
A;Accession: S46156
A;Molecule type: DNA
A;Residues: 1005-1916 <AIG>
A;Cross-references: EMBL:Z36144; MIPS:YBR275c
A;Experimental source: strain S288C
R;Holmstrom, K.; Brandt, T.; Kallesoe, T.
Yeast 10(Suppl.A), S47-S62, 1994
A;Title: The sequence of a 32420 bp segment located on the right arm of chromosome II fr
A;Reference number: S44537; MUID:94378722; PMID:8091861
A;Accession: S44537
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-1096 <HOL>
A;Cross-references: EMBL:X76053; NID:g600025; PID:g429120
A;Experimental source: strain S288C
R;Hardy, C.F.J.; Sussel, L.; Shore, D.
Genes Dev. 6, 801-814, 1992
A;Title: A RAP1-interacting protein involved in transcriptional silencing and telomer
A;Reference number: S25704; MUID:92249772; PMID:1577274
A;Accession: S25704
A;Molecule type: DNA
A;Residues: 1-579,'S',581-731,'A',733-1916 <HAR>
A;Cross-references: EMBL:X66501; NID:g4327; PIDN:CAA47121.1; PID:g4328
C;Genetics:
A;Gene: SGD:RIF1
A;Cross-references: SGD:S0000479; MIPS:YBR275c
A;Map position: 2R
C;Superfamily: Saccharomyces cerevisiae RIF1 protein
C;Keywords: transmembrane protein
F;263-279/Domain: transmembrane #status predicted <TM1>
F;296-312/Domain: transmembrane #status predicted <TM2>
F;438-454/Domain: transmembrane #status predicted <TM3>
F;663-679/Domain: transmembrane #status predicted <TM4>
F;788-804/Domain: transmembrane #status predicted <TM5>
F;905-921/Domain: transmembrane #status predicted <TM6>

  Query Match             3.2%;  Score 227;  DB 2;  Length 1916;
  Best Local Similarity   17.4%;  Pred. No. 0.0061;
  Matches  316;  Conservative  267;  Mismatches  612;  Indels  624;  Gaps   80;
```

```
Qy          3 HSKTRTNDGKITY-----------PPGVKEISDKI----------SKEEMVRRLKMV--- 38
Db        183 YKETRVNPRALEYWVSGEIHGLVDNESVSEFKEIIEGGLGILRQESEDYVARRFEVYATF 242

Qy         39 -----VKTFMDMDQDSEEEKELYLNLA---------LHLASDFFLKHPGKDVRLLVACCL 84
Db        243 NNIIPILTTKNVNEVDQKFNILIVNIESIIEICIPHLQIAQDTLLSSSEKKNPFVI---- 298

Qy         85 ADIFRIYAPEAPYTS--------------PD---KLKDIF-------------MFITR 112
Db        299 ----RLYVQIVRFFSAIMSNFKIVKWLTKRPDLVNKLKVIYRWTTGALRNENSNKIIITA 354

Qy        113 QLKGLEDTKSPQFNRYFYLLENI-AWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQ 171
Db        355 QVSFLRDEK---FGTFFLSNEEIKPIISTFTEIMEINSHNLIYEKL------LLIRGFLS 405

Qy        172 KVHMHMVDLMSS---------IICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALL 222
Db        406 KYPKLMIETVTSWLPGEVLPRIIIGDEIYSMKILITSIVVLL--------------ELL 450

Qy        223 KRTAQAIEPYITTFFNQVLMLGKTSISDLSEHVFDLILELY---NID----SHLLLSVLP 275
Db        451 KKCLDFVDEHERIY--QCIMLSPVCETIPEKFLSKLPLNSYDSANLDKVTIGHLLTQQIK 508

Qy        276 QLEFKLKSNDNEERLQV-VKLLAKMF--GAKDSELASQNKPLW-----QCYLGRFNDIHV 327
Db        509 --NYIVVKNDNKIAMDLWLSMTGLLYDSGKRVYDLTSESNKVWFDLNNLCFINN----HP 562

Qy        328 PIRLECVK---FASHCL-----MNHPDLAKDLTEYLK---------VRSHDPEEAIRHDV 370
Db        563 KTRLMSIKVWRIITYCICTKISQKNQEGNKSLLSLLRTPFQMTLPYVNDPSAREGIIYHL 622

Qy        371 IVSIVTA--AKKDIL-----LVNDHLLNFVRERTLDKRWRVR--------------KEA 408
Db        623 LGVVYTAFTSNKNLSTDMFELFWDHLITPIYEDYVFKYDSIHLQNVLFTVLHLLIGGKNA 682

Qy        409 MMGLAQIYKKY-------ALQSAAGKDAAK---QIAWIK---DKLLHIYYQNSIDDRLLV 455
Db        683 DVALERKYKKHIHPMSVIASEGVKLKDISSLPPQI--IKREYDKIMKVVFQT-------V 733

Qy        456 ERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQ 515
Db        734 EVAISNVNLAHDLILTS----------------------------LKHLPED------- 757

Qy        516 PKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVL-----------EDDEKIRKQLE 564
Db        758 RKDQTHLESFSSLILKVTQNNKDTPIFRDFFGAVTSSFVYTFLDLFLRKNDSSLVNFNIQ 817
```

```
Qy     565 VLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLL---------------ERIA 609
Db     818 ISKVGISQGNMTLDLLKDVIRKARN---ETSEFLIIEKFLELDDKKTEVYAQNWVGSTLL 874
Qy     610 PVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAET 669
Db     875 PPNISFREFQSLANIVNK-----------VPNENSIENFLDLCLKLSF--PVNL----- 915
Qy     670 FESLLACLKMDDEKVAEAALQIFKNTGSKIEEDF----------PHIRSALLPVLHHKS 718
Db     916 FTLLHVSMWSNNNFIYFIQSYVSKNE-NKLNVDLITLLKTSLPGNPELFSGLLPFL---- 970
Qy     719 KKGPPRQAKY---AIHCIHA--------------------------------IFSSKE- 741
Db     971 -----RRNKFMDILEYCIHSNPNLLNSIPDLNSDLLLKLLPRSRASYFAANIKLFKCSEQ 1025
Qy     742 ----------------TQFAQIFEPLHKSLDPSNLE--HLITPLVTIGHIALLAPDQFA 782
Db    1026 LTLVRWLLKGQQLEQLNQNFSEIENVLQNASD-SELEKSEIIRELL---HLAMANPIE-- 1079
Qy     783 APWKSWVATFIVKDLLMN--DRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMK 840
Db    1080 -PLFSGLLNFCIKNNMADHLDEFCGNMTSEVLF---KISPELLLKLLTYK--------- 1125
Qy     841 NNHSKSGTSTLRLLTTILHSDGD----LTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYH 896
Db    1126 --EKPNGKLLAAVIEKIENGDDDYILELLEKIIIQKE---------IQILEKLKEP--- 1170
Qy     897 EIITLEQYQLCALAINDECYQ-----VRQVFAQKLHKGLSRLRLPLEYMAICALCAKDP- 950
Db    1171 ----LLVFFLNPVSSNMQKHKKSTNMLRELVLLYLTKPLSRSAAKKFFSMLISILPPNPN 1226
Qy     951 -------------------VKERRAHARQCLVKNINVRREYLKQ----HAAVSEKLLSL 986
Db    1227 YQTIDMVNLLIDLIKSHNRKFKDKRTY--NATLKTIG---KWIQESGVVHQGDSSKEIEA 1281
Qy     987 LPE----YV--------------------VPYTIHLLAHDPDYVKVQDIEQLKDVKEC 1020
Db    1282 IPDTKSMYIPCEGSENKLSNLQRKVDSQDIQVPAT-QGMKEPPSSIQISSQISAKD-SDS 1339
Qy    1021 LWFVLEILMAKNENNSHA----------------FIRKMVENIKQTK----------D 1052
Db    1340 ISLKNTAIMNSSQQESHANRSRSIDDETLEEVDNESIREIDQQMKSTQLDKNVANHSNIC 1399
Qy    1053 AQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYL 1112
Db    1400 STKSDEVDVTE-LHESIDTQSSEVNAYQPIEVLTSELKAVTNRSIKTNPDHNVVNSDNPL 1458
Qy    1113 PPEMKSFFTPGKPKT------------------------TNVLGAVNKPLSSA------ 1141
Db    1459 KRPSKETPTSENKRSKGHETMVDVLVSEEQAVSPSSDVICTNIKSIANEESSLALRNSIK 1518
Qy    1142 ----------------GKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSE------- 1178
Db    1519 VETNCNENSLNVTLDLDQQTITKEDGKGQVEHVQRQENQESMNKINSKSFTQDNIAQYKS 1578
Qy    1179 ------MDHSENEDYTMSSPLPGKKSDKRDDSDLVRSE-----LEKPRGRKKTPVTEQEE 1227
Db    1579 VKKARPNNEGENNDYACNV----------EQASPVRNEVPGDGIQIPSGTILLNSSKQTE 1628
Qy    1228 KLGMDDLTK-------LVQEQKPKGSQRSRKRGHTASESDEQQWPEEKRLKEDILEN--- 1277
Db    1629 KSKVDDLRSDEDEHGTVAQEKHQVGAINSRNKNNDRMDSTPIQGTEEESREVVMTEEGIN 1688
Qy    1278 ---EDEQNSPPKKGKRGRPPKPLGGG-------------TPKEEPTMKTSKKGSKKKSG 1320
Db    1689 VRLEDSGTCELNKNLKG----PLKGDKDANINDDFVPVEENVRDEGFLKSMEHAVSKETG 1744
Qy    1321 PPAPEEEEEEE-----------------RQSGNTEQKSKSKQHRVSRRAQQRAESPESS 1362
Db    1745 -----LEEQPEVADISVLPEIRIPIFNSLKMQGSKSQIKEKLKKRLQRNELMPPDSPPRM 1799
Qy    1363 AIESTQSTPQKGRGRPSKT 1381
Db    1800 T-ENTNINAQNGLDTVPKT 1817
```

RESULT 14
T41581
hypothetical coiled-coil protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
C;Accession: T41581
R;Murphy, L.; Harris, D.; Wood, V.; Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z22002
A;Accession: T41581
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4717 <MUR>
A;Cross-references: EMBL:AL031546; PIDN:CAA20864.1; GSPDB:GN00068; SPDB:SPCC737.08
A;Experimental source: strain 972h(-)
C;Genetics:
A;Gene: SPDB:SPCC737.08
A;Map position: 3
C;Superfamily: Saccharomyces probable membrane protein YLR106c

```
  Query Match             3.1%;  Score 223.5;  DB 2;  Length 4717;
  Best Local Similarity   18.4%;  Pred. No. 0.03;
  Matches  290;  Conservative  249;  Mismatches  572;  Indels  469;  Gaps   71;

Qy      24 DKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLALHLASDFFLKHPGKDVRLLVACC 83
Db    2935 DDVS--NLIEQIKV-------LDLNDSIRFEIYLFLASRLCSE---KQHSSDTH-----S 2977
Qy      84 LADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVK---S 140
Db    2978 LANSFVLLANEF-YIHNAKIKQ------KELEEIEEKNRLYRQREFNFDKN-DYLKVFIN 3029
Qy     141 YNICFELEDSNEI------FTQLYRTLFSVINNGHNQKVHMHMVDLMSSI---------I 185
Db    3030 YDDEVEPEVEPEVVIERKRFLQLQFAFWSLYNEIYSEKMNVIPLEQLMNTGSYLAKKIKV 3089
Qy     186 CEGDTVSQELLDTVLVNLVPAHKNLNKQAY--------------------DLAKALLKR 224
Db    3090 KNPDMIASSGFDIVSVVLMMGVKSTNERQYWTPPVYNFYSDPNPSKAIEVRDLIKIVESR 3149
Qy     225 TAQAIEPYITTFFNQVLMLGKTSISDLS------------EHVFDLILEL-------YNI 265
Db    3150 AISLIKNWPENFVLRGLKDAIDAILNLSPFSPIAEYLSKLERVFHLLSEWEKLASREYSL 3209
Qy     266 DSHLLL---SVLPQLEFKLKSNDNEERLQVVKLLAKMFGA--------------KDSELA 308
Db    3210 ANEMDLIKKKIIDWRKFELSNWNNLLKLEEYKLSERVYPRLYSILQFIILKPFFENSKFT 3269
Qy     309 SQNKPLWQCYLGRF-NDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIR 367
Db    3270 KQNLCESASIIVQFITDLTVGEFQLCLK----CLLSFSQHAASL-------------RIC 3312
Qy     368 HDVIVSIVTAAKKDILLVN-------------------DHLLNFVRERTLDKRWRVRKE 407
Db    3313 HGI---------DAMLLNIYHYFEQFLSKVSEAIHTQKQSLENSIKERILLMSWKDT-- 3360
Qy     408 AMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHN 467
Db    3361 ---------NVYALKESAKKSHAELF-----KVLHRYRE------VLRQPVSSYLSQKHD 3400
Qy     468 LETTERMKCLYYLYATLDLNAVKALNEMW--KCQNLLRHQVK----DLLDLIKQPKTDAS 521
Db    3401 WD-----------SLLDTENNSA---MWVAKKVNLSPSYIEKMDTEIMKLV--PVRFSN 3443
Qy     522 VKAIFSKVMVITRNLPDPGK------------AQDFMKKFTQVLEDDE-------KIRKQ 562
Db    3444 TPTTLRLMWTLFANVEKPGSTFTNMVSNLITDARELMKLTPETINDDNLSEIKHLKSRKH 3503
```

```
Qy    563 LEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESISALI 622
Db   3504 L--LLTETFKTLKAFGLQYRV--KAGIEENLSN-----LRNLLAVIPTFPVTSLSIEKVD 3554

Qy    623 KQVNKSID-------GTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAE-----T 669
Db   3555 RSLMKSLDFIPKFQTLAGHQHNDLSVP---EVQKGVGL-----FNSMLSLQLGERAQLVE 3606

Qy    670 FESLLACLK--MDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAK 727
Db   3607 FTNELLALKNVYSEVGVNGSPLESFNNS------SFNEVSS--LGYDHDFENRA---QAV 3655

Qy    728 YAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPWKS 787
Db   3656 SMLCQIYAIVIQK----------HSSISPT------ASFQSIGH----ELSRFA----- 3689

Qy    788 WVATFIVKDLLMNDRLPG---------------------------KKTTKLWVPDEEV 818
Db   3690 --------DLLSNKLFPSSIPLYASADKVSSIRDQQKGINDLIEYCRKKRTEL--PELSY 3739

Qy    819 SPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRL 878
Db   3740 CFKHLVSLQSLKSISRTQVDLTND------EFLNLMNFVLNLFDSLLSSIETATKNMRTF 3793

Qy    879 R-LAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLP- 936
Db   3794 KELAETSSFIEMSS--CFSKVLRAFNLKFQSMKLSSLKEKLRSSSVDKMSCQLLMLFLPV 3851

Qy    937 -----------LEY--------------------MAICALCAKDPVKERRAHARQCLVK 964
Db   3852 CEQFINLAESVLDYFINVHNSNLDSLSKISTLFFMVANNGFCSPDLPQEGKSNSGE---- 3907

Qy    965 NINVRREYLKQHAAV-----SEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIE--QLKDV 1017
Db   3908 --------LESGTGLGSGVGAEDITNTLNE---DDDLEELANEEDTANQSDLDESEAREL 3956

Qy   1018 KECLWFVL-EILMAKNENNSHAFIRKMVENIKQTKDAQG---PDDA--KMNEKLYTVCDV 1071
Db   3957 ESDMNGVTKDSVVSENENSDS-------EEENQDLDEEVNDIPEDLSNSLNEKLW----- 4004

Qy   1072 AMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVL 1131
Db   4005 --------------DEP-----------------NEEDLLETEQKS---------NEQ 4022

Qy   1132 GAVNKPLSSAGKQSQTKS----SRMETVSNASSSSNPSSPGRIKGRL-DSSEMDHSENED 1186
Db   4023 SAANNESDLVSKEDDNKALEDKDRQEKEDEEEMSDDVGIDDEIQPDIQENNSQPPPENED 4082

Qy   1187 YTMSSP----LPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQK 1242
Db   4083 H-LDLPEDLKLDEKEGDVSKDSDLEDMDMEAADENKEEADAEKDEP--MQDFEDPLEENN 4139

Qy   1243 PKGSQRSRKRGHTASESDEQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTP 1302
Db   4140 TLDEDIQQDDFSDLAEDDEKM--NEDGFEENVQENEESTEDGVKSDEELEQ-----GEVP 4192

Qy   1303 KEEPTMKTSKKGSKKKSGPPAPEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESPESS 1362
Db   4193 EDQAIDNHPKMDAKSTFASAEADEENTDKGIVGENEELGEEDGAAESGVRGNGTADGEFS 4252

Qy   1363 AIESTQSTPQKGRGRPSKTP 1382
Db   4253 SAEQVQ------KGEDTSTP 4266
```

RESULT 15
T18296
myosin heavy chain - Entamoeba histolytica
C;Species: Entamoeba histolytica
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C;Accession: T18296
R;Guillen, N.
submitted to the EMBL Data Library, February 1997
A;Reference number: Z18865
A;Accession: T18296
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2139 <GUI>
A;Cross-references: EMBL:L03534; NID:g1850912; PID:g1850913; PIDN:AAB48065.1
C;Genetics:
A;Gene: mhcA
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;91-780/Domain: myosin motor domain homology <MMO>

```
  Query Match             3.1%;  Score 223;  DB 2;  Length 2139;
  Best Local Similarity   19.6%;  Pred. No. 0.011;
  Matches  292;  Conservative  233;  Mismatches  570;  Indels  396;  Gaps   62;

Qy      2 AHSKTRTNDGKI-TYPPGVKEISDKISKEEMVRR-LKMVVKTFMDMDQDSEEEKELYLNL 59
Db    894 ATAKTGELEAKIQDLEDKISELESKLSAAELDKQELNLKIE---NLEEDKEELKETIDNL 950

Qy     60 ALHLASDFFLKHPGKDVRLLVACCLADIFRIYAP--------EAPYTSPDKLKDIFMFIT 111
Db    951 KGDLKDS---KLKGEDLEVEITELNSQINTLNATVDKKDKTIAEMQESIDEKEDEITKLK 1007

Qy    112 RQLKGLEDTKSPQFNRYFYLLENIAWV---------KSYNICFELEDSNEIFTQLYRTLF 162
Db   1008 GDIKLLEEEKDD-------LEQDRADVSATKDDIAKKLNKITIECEDAKDEIAKLEQELE 1060

Qy    163 SVINNGHN-----QKVHMHMVDLMSSIIC----------EGDTVSQELLDTVLVNLVPAH 207
Db   1061 DEENKNKDLTNELQQTQLKLGETEKSLAAQVAATKKASDERDTLSQNLENEKLTT----- 1115

Qy    208 KNLNKQAYDLAKALLKRTAQAIEPYITTFFNQVLMLGKTSISDLSEHVFDLILELYNIDS 267
Db   1116 KNLTKTKADLEK--------------------------KISGLKQDYEDL-----EDDK 1143

Qy    268 HLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHV 327
Db   1144 NKIEGDLRNAQRKIKELDDE----ITK------GADVSQYLQKQKEEYESQIAKM----- 1188

Qy    328 PIRLECVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVND 387
Db   1189 ------------------------------QEEKEAIGNDVKNKEKTIKEKELEIQS- 1215

Qy    388 HLLNFVRERTLDKRWRVRKEAMMGLAQIYKKY-ALQSAAGKDAAKQIAWIKDKLLHIYYQ 446
Db   1216 ------LQEKLDETEVEKEDAEKKKKEIEKEMKALQEEKENVESSKNSTEKDK------- 1262

Qy    447 NSIDDRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQV 506
Db   1263 KKLED--------------NLKDTQ----------------KKLDDMTADNEKLKAKA 1290

Qy    507 KDLLDLIKQPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVL 566
Db   1291 KDLEAQLNEVQ-DNHEKAV-------------ADAELLNKKKAQSDKELNSLKAELEAL 1335

Qy    567 VSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQVN 626
Db   1336 T-------KAKSVVESKNKDSENEKAALS---EEIDQANEKLKNIQADLRKATADLQEAN 1385

Qy    627 KSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAE 686
Db   1386 EKKAEVEAQRDKLVADNKKMTKTLEEIKARDEENT---YKVENYEKVLKRKEADLEE-AN 1441

Qy    687 AALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHA----IFSSKET 742
Db   1442 ENLDIEKKDRMNKEKQV------------KKLEGELKETKDKLNAAIAEKDSIFTAKKQ 1488

Qy    743 QFAQIFEPLHKSLDPSNLEH--LITPLVTIGHIALLAPDQFAAPWKSWVATFIVKDLLMN 800
Db   1489 SDADL-EELNKTVE----EHDEVVAKLNT--QITKLTRDNQSAEEE-------LNELRSK 1534

Qy    801 DRLPGKKTTKLWVPDEEV-----------SPETMVKI---------QAIKMMVRWLLGM 839
```

```
Db       1535 ADKDKKKISEL---EEQVNELESRPVGTGNADENEIKIRDAQIADLNKALEMK-----GV 1586

Qy        840 KNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLA----------AGSAIVK- 888
              :||  ::    |:         | ||| : :|:: :| :|  |          |  |: :
Db       1587 QNNQLQATNKELKA------KDNDLTSKIEITENEMKKLENAKKRLEQDKDEADKAVSEQ 1640

Qy        889 -LAQEPCYHEIITL-EQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALC 946
               : ::     |:  |  :  |      ||       ::    |:|    :: |:  ||
Db       1641 TIKRKGLEEEVKKLTTEIQALKFQINAPSSVAQEEEKQRLESDIAELKEQLE-------- 1692

Qy        947 AKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYV 1006
                   :||   |        |   |   |  |  : |    :|  | |      |:| : :
Db       1693 -----QERTTAA--------NAEAERKKIQAELDEVKFNL--EDVTNQREKLVAKNSE-- 1735

Qy       1007 KVQDIEQLKDVKECLWFVLEILMAKNENNSHAF--IRKMVENIKQTKDAQGPDDAKMNEK 1064
                  :|: ||: |: |     :| :     |    |       : :     :  :||:    |  | ||
Db       1736 NDAEIDSLKEEKKALEDEIEKITDDNNKLSEEIDSLDRKYNALLDSKDS----DVSMKEK 1791

Qy       1065 LYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGK 1124
                               | |  :||: |                 || :  |         :|
Db       1792 FQDEL---------KVTKDALETEK-------------KNHAETM-----RLKGRLEKEA 1824

Qy       1125 PKTTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSEN 1184
                :     | |: | |  | :| : |:::     ::     |   :      :| :|| :: | ::
Db       1825 AEVQVRLEALQKNLDLA-QQEKAKATKDYRAADGELKSLMNELDDVKDQLDKAQDDLADK 1883

Qy       1185 EDYTMS-----SPLPGKKS--DKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKL 1237
              ||    :             |  :||  | |         : :|||  || |       |::|    :    :
Db       1884 EDELATLDQKYKTLVKQKSVFDSRIQEMQEQLDLEK-AGRAKA----QKQKQAYEKKLQE 1938

Qy       1238 VQE-----QKPKGSQRSRKRGHTASESDEQQWPEEKR-LKEDILENEDEQNSPPKKGKRG 1291
              :||       :: |  :     |      :| :  | |:  |::| ||:|      :: |      |   :
Db       1939 LQENDNDFEEYKETADKRINTLSAQKDDLQKELEKERGLKQD--SEKEVQRLRVKCQELE 1996

Qy       1292 RPPKPLGGG------------------TPKEEPTMKTSKKGSKK-KSGPPAPEE------ 1326
                      :||                 | :  |   :|    |   || |:        :|
Db       1997 TKVAEVGGANVSIAKVKAKYEAEIEELTTEAEDALKAKMKAEKKAKTSQKKLDELQKTIA 2056

Qy       1327 EEEEERQSGNTE----QKSKSKQHRVSRRAQQRAESPESSAIESTQSTPQK 1373
              : |  :  | |||     |      |   :   |   :  |  | |     :  |  :    |
Db       2057 DYETKEASFNTEIGKTQAELKKYQQQVRDDETRMSSLEDEIKKGTDALANK 2107
```

Search completed: September 24, 2003, 20:21:28
Job time : 42 secs

```
                 GenCore version 5.1.6
        Copyright (c) 1993 - 2003  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         September 24, 2003, 20:20:49 ; Search time 236 Seconds
                                           (without alignments)
                                           891.814 Million cell updates/sec

Title:          US-09-512-581B-2
Perfect score:  7193
Sequence:       1 MAHSKTRTNDGKITYPPGVK..........QKGRGRPSKTPSPSQPKKNV 1391

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       566894 seqs, 151307093 residues

Total number of hits satisfying chosen parameters:      566894

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_AA:*
                1:  /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
                2:  /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
                3:  /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
                4:  /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
                5:  /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
                6:  /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
                7:  /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
                8:  /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
                9:  /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
                10:  /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
                11:  /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
                12:  /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
                13:  /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
                14:  /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
                15:  /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
                16:  /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
                17:  /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
                18:  /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	758.5	10.5	363	9	US-09-925-302-800	Sequence 800, App
2	231.5	3.2	2835	10	US-09-885-535-4	Sequence 4, Appli
3	225	3.1	1743	12	US-09-882-227-624	Sequence 624, App
4	206	2.9	2243	15	US-10-118-513A-12	Sequence 12, Appl
5	197.5	2.7	1530	15	US-10-118-513A-6	Sequence 6, Appli
6	197.5	2.7	2266	15	US-10-118-513A-14	Sequence 14, Appl
7	189	2.6	3907	15	US-10-171-311-2	Sequence 2, Appli
8	189	2.6	3925	15	US-10-171-311-6	Sequence 6, Appli
9	186.5	2.6	2649	12	US-10-205-219-169	Sequence 169, App
10	186	2.6	1979	15	US-10-205-823-419	Sequence 419, App
11	185.5	2.6	2437	9	US-09-815-242-5834	Sequence 5834, Ap
12	185.5	2.6	6281	9	US-09-815-242-12996	Sequence 12996, A
13	185	2.6	780	10	US-09-770-689A-5	Sequence 5, Appli
14	184.5	2.6	1814	15	US-10-128-714-3388	Sequence 3388, Ap
15	184	2.6	676	11	US-09-823-187-24	Sequence 24, Appl
16	183.5	2.6	1400	10	US-09-764-176-7	Sequence 7, Appli
17	183	2.5	670	11	US-09-823-187-86	Sequence 86, Appl
18	183	2.5	671	11	US-09-946-374-308	Sequence 308, App
19	183	2.5	671	11	US-09-823-187-85	Sequence 85, Appl
20	183	2.5	671	12	US-10-015-387A-308	Sequence 308, App
21	183	2.5	671	12	US-10-006-130A-308	Sequence 308, App
22	183	2.5	671	12	US-10-199-672-346	Sequence 346, App
23	183	2.5	671	12	US-10-006-172A-308	Sequence 308, App
24	183	2.5	671	12	US-10-187-749-346	Sequence 346, App
25	183	2.5	671	12	US-10-194-457-346	Sequence 346, App
26	183	2.5	671	12	US-10-184-642-346	Sequence 346, App
27	183	2.5	671	12	US-10-196-747-346	Sequence 346, App
28	183	2.5	671	12	US-10-015-392A-308	Sequence 308, App
29	183	2.5	671	12	US-10-017-253A-308	Sequence 308, App
30	183	2.5	671	12	US-10-173-689-346	Sequence 346, App
31	183	2.5	671	12	US-10-173-690-346	Sequence 346, App
32	183	2.5	671	12	US-10-173-691-346	Sequence 346, App
33	183	2.5	671	12	US-10-173-692-346	Sequence 346, App
34	183	2.5	671	12	US-10-173-694-346	Sequence 346, App
35	183	2.5	671	12	US-10-173-698-346	Sequence 346, App
36	183	2.5	671	12	US-10-173-699-346	Sequence 346, App
37	183	2.5	671	12	US-10-173-707-346	Sequence 346, App
38	183	2.5	671	12	US-10-174-569-346	Sequence 346, App
39	183	2.5	671	12	US-10-174-583-346	Sequence 346, App
40	183	2.5	671	12	US-10-174-587-346	Sequence 346, App
41	183	2.5	671	12	US-10-174-589-346	Sequence 346, App
42	183	2.5	671	12	US-10-174-591-346	Sequence 346, App
43	183	2.5	671	12	US-10-175-736-346	Sequence 346, App
44	183	2.5	671	12	US-10-175-742-346	Sequence 346, App
45	183	2.5	671	12	US-10-175-744-346	Sequence 346, App

ALIGNMENTS

```
RESULT 1
US-09-925-302-800
; Sequence 800, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA104
;  CURRENT APPLICATION NUMBER: US/09/925,302
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05918
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 896
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 800
;   LENGTH: 363
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (358)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-800

  Query Match             10.5%;  Score 758.5;  DB 9;  Length 363;
  Best Local Similarity   45.0%;  Pred. No. 5e-42;
  Matches  172;  Conservative   56;  Mismatches 105;  Indels   49;  Gaps   11;

Qy        965 NINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFV 1024
              ||::||||:||:   :|||||||||||||| |||||||||: : ||::||:|:||||||:
Db          1 NISIRREYIKQNPMATEKLLSLLPEYVVPYMIHLLAHDPDFTRSQDVDQLRDIKECLWFM 60

Qy       1025 LEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYS 1084
              ||:|| |||||||||::|| |||| |:||| ||::| |||||||||||: :| |||   :
Db         61 LEVLMTKNENNSHAFMKKMAENIKLTRDAQSPDESKTNEKLYTVCDVALCVINSKSALCN 120
```

```
Qy       1085 LESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQ 1144
               :|||||||| :|||||:|:| | |:|: |  :        ||||   ||||||||||: |::
Db        121 ADSPKDPVLPMKFFTQPEKDFCNDKSYISEETRVLLLTGKPKPAGVLGAVNKPLSATGRK 180

Qy       1145 SQTKSSRMETVS--NASSSSNPSSPGRIKGR-LDSSEMDHSENED--YTMSSPLPGKKSD 1199
               :|:  || |   | :|  |||:  | : :  :::|    ||||:    : |  | |  |
Db        181 PYVRSTGTETGSNINVNSELNPSTGNRSREQSSEAAETGVSENEENPVRIISVTPVKNID 240

Qy       1200 KRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASES 1259
                                    || :|  :  |  |     |     | | :|| ||: :
Db        241 --------------------PVKNKE--INSDQAT-----QGNISSDRGKKRTVTAAGA 272

Qy       1260 DE-QQWPEEKRLKEDILENEDEQNSP-PKKGKRGRPPKPLG-GGTPKEEPTMKTSKKGSK 1316
               :  ||  :||            ||   | | | :||| ||    |    | :    | || |
Db        273 ENIQQKTDEK---------VDESGPPAPSKPRRGRRPKSESQGNATKNDDLNKPINKGRK 323

Qy       1317 KKSGPPAPEEEEEEERQSGNTE 1338
               :    |  :|     ::|| :
Db        324 R----AAVGQESPGGLEAGNAK 341
```

RESULT 2
US-09-885-535-4
; Sequence 4, Application US/09885535
; Patent No. US20020104105A1
; GENERAL INFORMATION:
; APPLICANT: Myriad Genetics, Inc.
; APPLICANT: Heichman, Karen
; APPLICANT: Bartel, Paul L.
; TITLE OF INVENTION: Protein-Protein Interactions
; FILE REFERENCE: 2318-266-II
; CURRENT APPLICATION NUMBER: US/09/885,535
; CURRENT FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/213,245
; PRIOR FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;   LENGTH: 2835
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-885-535-4

```
  Query Match             3.2%;  Score 231.5;  DB 10;  Length 2835;
  Best Local Similarity   18.1%;  Pred. No. 6.1e-06;
  Matches  323;  Conservative  272;  Mismatches  635;  Indels  555;  Gaps   72;

Qy         22 ISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLALHLASDFFLKHPGKDVRLLVA 81
               : ::  | |: |:|: ::|:| |::| :|          |    |||  |        |
Db        168 VKEEYLKAELSRQLEGILKSFKDVEQKAE-------NHVQHLQSACASSHQ--------- 211

Qy         82 CCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSY 141
                                          |  ::|  :   |||  :  |           ||:
Db        212 ------------------------FQQMSRDFQAWLDTKKEEQN-----------KSH 234

Qy        142 NICFELEDSNEIFTQLYRTLFSVINNGHN------QKVHMHMVDLMSSIICEGDTV---- 191
               |  :|:                     | |:|  :          |  ||:  :
Db        235 PISAKLD----------VLESLIKDHKDFSKTLTAQSHMY-----EKTIAEGENLLLKT 278

Qy        192 -------SQELLDTVLVNLVPAHKNL----NKQAYDLAKAL-LKRTAQAIEPYITTFFNQ 239
                     |  |:|:  |    :| :     ||    | |||  |    :  : |:|      |
Db        279 QGSEKAALQLQLNTIKTNWDTFNKQVKERENKLKESLEKALKYKEQVETLWPWIDKCQNN 338

Qy        240 VLML--------GKTSIS-------DLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSN 284
               :  :          |: ||:         :: :|  |  :: || |   :: ||||  ::: :: ::
Db        339 LEEIKFCLDPAEGENSIAKLKSLQKEMDQH-FGMV-ELLNNTANSLLSVC-EIDKEVVTD 395

Qy        285 DNEERLQVVKLLAKMFGAKD------------SELASQNKPLWQCYLGRFNDIHVPI-- 329
               :|:  :| | ::  :    :|             |:: ::|   ||     :   ||| :
Db        396 ENKSLIQKVDMVTEQLHSKKFCLENMTQKFKEFQEVSKESKRQLQCAKEQL-DIHDSLGS 454

Qy        330 ---------RLECVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKK 380
                        |:  :  :    |  :  ||||  |  :  |  |  :  |  :        ||:: :  |  |::

Db        455 QAYSNKYLTMLQTQQKSLQALKHQVDLAKRLAQDLVVEASDSKGT--SDVLLQVETIAQE 512

Qy        381 DILLVND---------------HLLNFVRE---------------------RTLDKRW 402
               |                   |  |  :||                      || |:
Db        513 HSTLSQQVDEKCSFLETKLQGIGHFQNTIREMFSQFAEFDDELDSMAPVGRDAETLQKQK 572

Qy        403 RVRK------EAMMGLAQIYKKYALQSAAGKDAAKQIAWIKD----------------- 438
               |        ||:|      |       | :: :  :  ||
Db        573 ETIKAFLKKLEALMASNDNANKTCKMMLATEETSPDLVGIKRDLEALSKQCNKLLDRAQA 632

Qy        439 ---------KLLHIYYQNSIDDRLLVERI----FAQYMVPHNLETTERMKCLYYLYATLD 485
                        | |  :|     :  :|:::      :|  |     ||    :: :: :    :
Db        633 REEQVEGTIKRLEEFYSKLKEFSILLQKAEEHEESQGPVGMETETINQQLNMFKVFQKEE 692

Qy        486 L------------------NAVKA------------LNEMWK-----------CQNLL 502
               :                   :| |:                  :|  ||         | |
Db        693 IEPLQGKQQDVNWLGQGLIQSAAKSTSTQGLEHDLDDVNARWKTLNKKVAQRAAQLQEAL 752

Qy        503 RH--QVKDLLDLI-------------KQPKTDAS-VKAIFSKVMVITRNLPDPGKAQDF 545
               |  : :| |: :                | |  :   |||   :   :: | | |     :
Db        753 LHCGRFQDALESLLSWMVDTEELVANQKPPSAEFKVVKAQIQEQKLLQRLLDDRKSTVEV 812

Qy        546 MKK------FTQVLEDDEKIRKQLEVLVS--------PTCSCKQAEGCVREITKKLGNPK 591
               :|:       |     |  || ||| :| |                 :| ||  :   : ::
Db        813 IKREGEKIATTAEPADKVKILKQLSLLDSRWEALLNKAETRNRQLEG-ISVVAQQFHETL 871

Qy        592 QPTNPFLEMIKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLE 651
               :| | :|  |:  |    |:        :   |  |  :|  :                 :  |  |||: |||
Db        872 EPLNEWLTTIEKRLVNCEPIGTQASKLEEQIAQ-HKVLQEDILLRKQNV--DQALLNGLE 928

Qy        652 LLKVLSFTHPISFHSAETFESL------LACLKMDDEKVAEAALQIFKNTGSKIEEDFPH 705
               |||  :         :      :  |::         |     |     |   | |||: :   |  ||
Db        929 LLKQTTGDEVLIIQ--DKLEAIKARYKDITKLSTDVAKTLEQALQLARRLHSTHEELCTW 986

Qy        706 IRSALLPVLHHKSK--KGP--------PRQAKYAIHCIHAIFSSKETQFAQIF------- 748
               :      : :| ::::  ||           |:: |           |:  |      : :
Db        987 LDKVEVELLSYETQVLKGEEASQAQMRPKELKKEAKNNKALLDSLNEVSSALLELVPWRA 1046

Qy        749 -EPLHKSLDPSN-----LEHLITPLVTIGHIALLAPDQF---AAPWKSWVATFIVKDLLM 799
               | | | :  |     :  ||  |       |:|  ||   |       ||:    | : :
Db       1047 REGLEKMVAEDNERYRLVSDTITQKVEEIDAAILRSQQFDQAADAELSWITETEKKLMSL 1106

Qy        800 ND---------RLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTS 849
               |         :|  :||  :  :    :  :   :  :||   :   |  ||:
Db       1107 GDIRLEQDQTSAQLQVQKTFTMEILRHKDIIDDLVK-SGHKIMTACSEEEKQSMKKKLDK 1165

Qy        850 TLRLLTTILHSDGDLTEQGKISKPDMSRL---------RLAAGSAIVKLAQEPCYHEIIT 900
               |:  ||  :  :   | : :: ::::                     |    :|:    |   |   |
Db       1166 VLKNYDTICQINSERYLQLERAQSLVNQFWETYEELWPWLTETQSIISQLPAPAL-EYET 1224

Qy        901 LEQYQLCALAINDECYQVRQVFAQ------KLHK-GLSRLRL-PLE-------YMAICAL 945
               | | |        :|  |:|:: |:       |::|  |   | | | |           |:|  |
Db       1225 LRQQQ-------EEHRQLRELIAEHKPHIDKMNKTGPQLLELSPGEGFSIQEKYVAADTL 1277

Qy        946 CA--KDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDP 1003
               :  |: ||:|     | : |:    ::   :    : | |  ::         | :|   : :
Db       1278 YSQIKEDVKKRAV----ALDEAISQSTQFHDKIDQILESLERIVERLRQPPSI---SAEV 1330

Qy       1004 DYVKVQ----------DIEQLKDVKECLWFVLEILMAK--------------NENNSHAFI 1040
               : :|  |                |:|:|: | |  |    | ::|:                    :   ||
Db       1331 EKIKEQISENKNVSVDMEKLQPLYETLKQRGEEMIARSGGTDKDISAKAVQDKLDQMVFI 1390

Qy       1041 RKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQ 1100
               : :  : ::|:  |  :: || :  ||   |:   |  |    :  :|| :     |
Db       1391 WENIHTLVEEREAKLLDVMELAEKFW--CDHMSLIVTIKDTQDFIRDLEDPGIDPSVVKQ 1448

Qy       1101 PDKNFSNTK---NYLPPEMKSFFT-----------PGKPKTTNVLGAVNKPLSSAGKQSQ 1146
               :     :  |  |:                    |||        :  :|      |   |  :
Db       1449 QQEAAETIREEIDGLQEELDIVINLGSELIAACGEPDKPIVKKSIDELNSAWDSLNKAWK 1508

Qy       1147 TKSSRMETVSNAS-------------------SSSNPSSPGRIKGRLDSSEMDHSENE 1185
               :  ::|   |:                            :|  :|            :   |:   ::|
Db       1509 DRIDKLEEAMQAAVQYQDGLQAVFDWVDIAGGKLASMSPIGTDLETVKQQIEELKQFKSE 1568
```

```
Qy       1186 DYTM---------SSPLPGKKSDKRDDSDLVRS---ELEKPRGRKKTPVTEQEEKL---- 1229

Db       1569 AYQQQIEMERLNHQAELLLKKVTEESDKHTVQDPLMELKLIWDSLEERIINRQHKLEGAL 1628

Qy       1230 --------GMDDLTK-------LVQEQKPKGS-------------------------QR 1248

Db       1629 LALGQFQHALDELLAWLTHTEGLLSEQKPVGGDPKAIEIELAKHHVLQNDVLAHQSTVEA 1688

Qy       1249 SRKRGHTASESD--------------EQQWPEEKRLKEDILENEDEQNSP------PKK 1287

Db       1689 VNKAGNDLIESSAGEEASNLQNKLEVLNQRW-------QNVLEKTEQRKQQLDGALRQAK 1741

Qy       1288 GKRGR---------------PPKPLGG--GTPKE--------------EPTMKT-SKK 1313

Db       1742 GFHGEIEDLQQWLTDTERHLLASKPLGGLPETAKEQLNVHMEVCAAFEAKEETYKSLMQK 1801

Qy       1314 GSKKKSGPPAPEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAES 1358

Db       1802 GQQMLARCP-KSAETNIDQDINNLKEKWESVETKLNERKTKLEEA 1845
```

RESULT 3
US-09-882-227-624
; Sequence 624, Application US/09882227
; Publication No. US20030158396A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Ooomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20030158396A1el Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/047002
; CURRENT APPLICATION NUMBER: US/09/882,227
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/902,615
; PRIOR FILING DATE: 1997-07-29
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 624
;   LENGTH: 1743
;   TYPE: PRT
;   ORGANISM: Helicobacter pylori
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: 876
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-09-882-227-624

```
  Query Match             3.1%;  Score 225;  DB 12;  Length 1743;
  Best Local Similarity   18.4%;  Pred. No. 8.2e-06;
  Matches  285;  Conservative  217;  Mismatches  523;  Indels  520;  Gaps   71;

Qy          4 SKTRTN-DGKITYPPGVKEISDKISKEEMVRRLK--------------------MVVKT 41

Db        312 TETQTNFDGDKS-----EEITDDSNDQEIIKGSKKKYIIGGIVVAVLIVIILFSRSIFHY 366

Qy         42 FMDMDQDS---EEEKELYLNLALHLASDF--FLKHPGKDVRLLVACCLADIFRIYAPEAP 96

Db        367 FMPLEDKSSRFSKDRNLYVNDEIQIRQEYNRLLKERNEKGNMID--------------- 410

Qy         97 YTSPDKLKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQ 156

Db        411 -------KNLFF----------NDDPNRTLYNYL--NIA---------EIEDKNPL--- 438

Qy        157 LYRTLFSVINNGHNQKVHMHMVDLMSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYD 216

Db        439 --RAFYECISNGGNYEECLKLI--------------------------KDKKLQDQ--- 466

Qy        217 LAKALLKRTAQAIEPYITTFFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQ 276

Db        467 -----MKKTLEA--------YNDCIKNAKT---------------------------- 483

Qy        277 LEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKF 336

Db        484 ---------EEERIKCLDLI------KDENL---KKSL-------LNQQKVQVALDCLKN 518

Qy        337 A------SHC--LMNHP----------DLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAA 378

Db        519 AKTDEERNECLKLINDPEIREKFRKELELQKELQEY------------KDCIKNAKTEA 565

Qy        379 KKDILL--VNDHLLNFVRERTLD-----KRWRVRKEAMMGLAQIYKKYALQ--------- 422

Db        566 EKNKCLKGLSKEAIERLKQQALDCLKNAKTDEERNECLKNIPQDLQKELLADMSVKAYKD 625

Qy        423 --SAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLET-TERMKCLYY 479

Db        626 CVSKARNEKEKQEC---EKLLTPEARKKLEQQVL--------DCLKNAKTDEERKKCLKD 674

Qy        480 LYATL--DLNAVKALN------------EMWKCQNLL-----------RHQVKDLLDL 512

Db        675 LPKDLQSDILAKESLKAYKDCVSQAKTEAEKKECEKLLTPEAKKLLEEEAKESVKAYLDC 734

Qy        513 IKQPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLED------DEKIRKQLEVL 566

Db        735 VSQAKTEAEKKEC-EKLLT-----PEAKKKLEEAKKSVKAYLDCVSRARNEKEKKECEKL 788

Qy        567 VSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQVN 626

Db        789 LTPEAK-KLLEQQALDCLKNAKTDKERKKCLKDLPKDLQKKV----LAKESVKAYLDCVS 843

Qy        627 KSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAE 686

Db        844 QAKTEAEKKECEKLLTPEA-RKLLEEAK------------KSVKAXLDC-------VSQ 882

Qy        687 AALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAI----HCIHAIFSSKET 742

Db        883 AKTEAEKKECEKLLT--PEARKLL-------------EEAKESVKAYKDCVSKARNEKEK 927

Qy        743 QFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPWKSWVATFIVKDLLMNDR 802

Db        928 K--------------ECEKLLTP-----EAKKLLEQQ------------VLDCLKNAK 954

Qy        803 LPGKKTTKLWVPD--EEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHS 860

Db        955 TEADK--KRCVKDLPKDLQKKVLAK-ESVKAYLDCVSRARNEKEKK------------- 997

Qy        861 DGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQ 920

Db        998 -----ECEKLLTPEAKKLLEEAKESL--KAYKDCLSQARNEEERRACEKLLTPE---ARK 1047

Qy        921 VFAQKLHKGL-------SRLRLPLEYMAICALCAKDPVKERRAHARQCLVK--NINVRRE 971

Db       1048 LLEQEVKKSIKAYLDCVSRARNEKEKKECEKLLTPEARKFLAKQVLNCLEKAGNEEERKA 1107

Qy        972 YLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAK 1031

Db       1108 CLKN-----------LPKDLQE---NILAK----------ESLKAYKDCL------SQAR 1137

Qy       1032 NENNSHAF-------IRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYS 1084

Db       1138 NEEERRACEKLLTPEARKLLE-----------QEVKKSVKAYLDCVSRARNEKEKKECEK 1186

Qy       1085 LESPKDPVLPARFFTQPDKNFSN-TKNYLP---------------PEMKSFFTPGKPKTT 1128

Db       1187 LLTPEARKFLAKELQQKDKAIKDCLKNADPNDRAAIMKCLDGLSDEEKLKYLQEAREKAV 1246

Qy       1129 NVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSE-MDHSEN--E 1185

Db       1247 ADCLAMAKTDEEKRKCQNLYSDLIQEIQNKRTQNKQNQLSKTERLHQASECLDNLDDPTD 1306

Qy       1186 DYTMSSPLPGKKSDKR----------DDSDLVRSELEKPR-----GRKKTPVTEQEEKLG 1230
```

```
Db       1307 QEAIEQCLEGLSDSERALILGIKRQADEVDLIYSDLRNRKTFDNMAAKGYPLLPMDFKNG 1366

Qy       1231 MD---------DLTKLVQEQKPKGSQRSRKRGHTASESD-EQQWPEEKRLKEDILENEDE 1280
Db       1367 GDIATINATNVDADKIASDNPI----------YASIEPDIAKQYETEKTIKDKNLE---- 1412

Qy       1281 QNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEERQSGNTEQK 1340
Db       1413 ----------AKLAKALGGN--KKDDDKEKSKKSTAEAKAENNKIDKDVAETAKNISEIA 1460

Qy       1341 SKSKQHRVSRRAQQRAE--SPESSAIESTQSTPQK----GRGRPS 1379
Db       1461 LKNKKEKSGEFVDENGNPIDDKKKAEKQDETSPVKQAFIGKSDPT 1505

RESULT 4
US-10-118-513A-12
; Sequence 12, Application US/10118513A
; Publication No. US20030039995A1
; GENERAL INFORMATION:
;  APPLICANT: Taga, Tetsuya
;  APPLICANT:  Kimura, Naoki
;  TITLE OF INVENTION: THE YS68 GENE INVOLVED IN PRIMITIVE HEMATOPOIESIS
;  FILE REFERENCE: 06501-107US1
;  CURRENT APPLICATION NUMBER: US/10/118,513A
;  CURRENT FILING DATE:  2002-04-08
;  PRIOR APPLICATION NUMBER: PCT/JP00/05756
;  PRIOR FILING DATE: 2000-08-25
;  PRIOR APPLICATION NUMBER: JP 11-288738
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: JP 11-288739
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: JP 2000-123721
;  PRIOR FILING DATE: 2000-04-19
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 2243
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-118-513A-12

  Query Match             2.9%;  Score 206;  DB 15;  Length 2243;
  Best Local Similarity   18.0%;  Pred. No. 0.00021;
  Matches  290;  Conservative  218;  Mismatches  509;  Indels  590;  Gaps   74;

Qy        195 LLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQA-IEPYITTF----------------- 236
Db        526 LLSPRLIDIQPSSLSQEEQLEAILSAAIQTSSLGLLTGYIRTWIIEEQPNSAANLRFVLE 585

Qy        237 --FNQVLMLGKTSISDLSEHVFD---LILELYNIDS----HLLLSVLPQLE--FKLKSN- 284
Db        586 WTWNKV-VLTKEEFDRLCVPLFDGSCRFIDPQTIQSIQQCHLLLSNLSTVLSCFAMEAQG 644

Qy        285 -------DNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFA 337
Db        645 ITERGLVDLSNKHMVTQLLCQY----------AHMVLWFCHSGL-----LPEGLDDALHL 689

Qy        338 SHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERT 397
Db        690 SRLRYNYPVIQNYYTSRRQKSERSPRGKWNHD-------------CLMIDGLVSQLGDE- 735

Qy        398 LDKRWRVRKEAMMG---------LAQIY--------KKYA---------LQSAAGK---- 427
Db        736 VEKLWK-RDEGGTGRYPPASIHALLDIYLLDNITEASKHAITIYLLLDIMYSFPNKTDTP 794

Qy        428 ----DAAKQIAWIKDKLL-------HIYYQNSIDDRLLVERIFAQ-------YMVPHNLE 469
Db        795 IESFPTAFAISWGQVKLVQGFWLLDHNDYENGLD--LLFHPVTAKPASWQHSKIIEAFMS 852

Qy        470 TTERMKCLYYLY-----------ATLDLNAV---KALNEMWKC--QNLLRHQVKDLL--- 510
Db        853 QGEHKQALRYLQTMKPTVSSSNEVILHLTVLLFNRCMVEAWNLLRQNSNRVNIEELLKHA 912

Qy        511 -----------DLIKQPKTDAS----VKAIFSKVMVITRNLPDPGKAQDFM--------- 546
Db        913 YEVCQEMGLMEDLLKLPFTNTEQECLVKFLQSSTSV---------ENHEFLLVHHLQRAN 963

Qy        547 ----KKFTQVLED------DEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNP 596
Db        964 YISALKLNQILKNNLMSDRDPRLRE-------------------RSVTR---------NS 995

Qy        597 FLEMI---------KFLLERIAPVHIDTESI----------SALIKQVNKSIDGTADDED 637
Db        996 ILDQYGKILPRVQRKLAVERAKPYHLSTSSVFHEVSRPKPLSAFPK---KAITGTVLTRS 1052

Qy        638 EGVP----------TDQAIRAGLELLK----------VLSFTHPISFHSAETFESLLACL 677
Db       1053 TFISNVLSKIGEVWASHEPRNGVSLFNSPKTEQPSPVVHSFPHP---------------- 1096

Qy        678 KMDDEKVAEAALQI-FKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIH-- 734
Db       1097 -----ELPEAFVGTPISNTSQRI--------SRLLDLVVH-----PVPQPSQCLEFIQQS 1138

Qy        735 ------AIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVT--IGHIALLAPDQFAAPWK 786
Db       1139 PTRSPLCLLSSSLPLSSQFKRPHQNTSRPSELLLLETPLIVKKAKSLALSAT-------S 1191

Qy        787 SWVATFIVKDLLMNDRLPGKKTTKL----WVPDEEVSPETMVKIQAIKMMVRWLLGMKNN 842
Db       1192 SGFAEFTPPSILRS----GFRTTPLASPSLSPGRSLTPPFRVKETRISFMEE---GMNTH 1244

Qy        843 HSKSGT---STLRLLTTILHSDG-------------------DL--------------TE 866
Db       1245 WTDRATDDRNTKAFVSTSFHKCGLPAETEWMKTSDKNTYFPLDVPAKGPQKVVAESLATH 1304

Qy        867 QGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCAL----------------A 910
Db       1305 SGRLEKLDVSKEDSTASTRSDQTSLE--YHDAPSPEDLEGAVFVSPKPASSSTELTTNST 1362

Qy        911 INDECYQVRQVF------------------AQKLHKGLSRLRLPLE----YMAICALC-- 946
Db       1363 LQTERDNDKDAFKSEGTPSPVKKQIGTGDAAVEAFSELSRLD-PVERAEASFGVSSVCEG 1421

Qy        947 -----------------------------------------------AKDPVKERRAHAR 959
Db       1422 ETSTSNSKTSVLDGIVPIESRTSILTADHKESVANTVADVESSGSTSSKCPVTSERSLGQ 1481

Qy        960 QCLVKNINVRREYLKQHAAVSEKLLSLLPE----YVVP---YTIHLLAHDPDYVKVQDIE 1012
Db       1482 KL---TLNLKEDEIEAH--VPKENVGLPEESPRISAAPSDTHEIHLIG--CENLEVQNSE 1534

Qy       1013 QLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVA 1072
Db       1535 E--EAKNLSFDELYPLGAEKLEYNLSTIEQQFCDLPDDKDSAECDAAEVDGELF------ 1586

Qy       1073 MNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLG 1132
Db       1587 ---VAQSNFTLILEGEEG----------EAEASDSAAPNMLPKSTK-------------- 1619

Qy       1133 AVNKPLSSAGKQSQTKSSRMETVSNASSSSN-----PSSPGRIKGRLDSSEMDHSENEDY 1187
Db       1620 --EKPVCYREPHNQERVTDLPSAVTADQESHKVETLPYVPEPVKVAIAENLLDVIKDTRS 1677

Qy       1188 TMSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKL------GMDDLT------ 1235
Db       1678 KEATPVAAGEAGDEDGAVIVSKAAHSSRLTNSTPKTVKEPRAETVNTSQSDDMVSSRTLT 1737

Qy       1236 --------KLVQEQKPKGSQRSRKRGHTASE---------SDEQQWPE-----------E 1267
Db       1738 RRQHALSLNVTSEQEPSAVATPKKRTRKIKETPESSERTCSDLKVAPENQLTAQNPPAPR 1797

Qy       1268 KRLKEDI----------LENEDE-QNSPPKKGKRGRPPKPLGGGTP-KEEPTMKTSKKGS 1315
Db       1798 RRKKKDVSQGTLPSSGAVEPEPEPQGTPGRLRLRTQPPEPAAEETPSRTKVRLSSVRKGT 1857
```

```
Qy       1316 KKK----------------SGPPA--------------PEEEEEEERQSGNTEQKSKS 1343
               ::                 |  |               : :  | :|  :::|:
Db       1858 PRRLKKSVENGQSIEILDDLKGSEAASHDGTVTELRNANLEDTQNMEYKQDEHSDQQPPL 1917

Qy       1344 KQHRVSRRAQQRAESPESSAIESTQ---------STPQKGRGRPSK 1380
               |: ||  |     :    |    ::|:|        :||:| |||| |
Db       1918 KRKRVREREVSVSSVTEEPKLDSSQLPLQTGLDVPATPRK-RGRPRK 1963


RESULT 5
US-10-118-513A-6
; Sequence 6, Application US/10118513A
; Publication No. US20030039995A1
; GENERAL INFORMATION:
;  APPLICANT: Taga, Tetsuya
;  APPLICANT:  Kimura, Naoki
;  TITLE OF INVENTION: THE YS68 GENE INVOLVED IN PRIMITIVE HEMATOPOIESIS
;  FILE REFERENCE: 06501-107US1
;  CURRENT APPLICATION NUMBER: US/10/118,513A
;  CURRENT FILING DATE:  2002-04-08
;  PRIOR APPLICATION NUMBER: PCT/JP00/05756
;  PRIOR FILING DATE: 2000-08-25
;  PRIOR APPLICATION NUMBER: JP 11-288738
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: JP 11-288739
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: JP 2000-123721
;  PRIOR FILING DATE: 2000-04-19
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 1530
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-118-513A-6

  Query Match             2.7%;  Score 197.5;  DB 15;  Length 1530;
  Best Local Similarity   18.2%;  Pred. No. 0.00044;
  Matches  293;  Conservative  223;  Mismatches  603;  Indels  493;  Gaps   68;

Qy         56 YLNLALHLASDFFL--------KHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIF 107
               |   :||   | :|          ||         :    | ||   :  |    :   :|
Db         14 YPPASLHAVLDMYLLDGVTEAAKHS------ITIYLLLDIMYSF-PNKTDTPIESFPTVF 66

Qy        108 MFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINN 167
                     |:| ::          |:|:::  :     :: |      :   :  |   : : : : ::
Db         67 AISWGQVKLIQG---------FWLIDHNDYESGLDLLFHPATAKPLSWQHSKIIQAFMSQ 117

Qy        168 GHNQKVHMHMVDLMSSIICEGDTVSQELLDTVLV---------NLVPAHKNLNKQAYDLA 218
               | :::  :  :  |    :  |: |   |  |||:                |  :  | |
Db        118 GEHRQA-LRYIQTMKPTVSSGNDVILHL--TVLLFNRCMVEAWNFLRQHCN--------- 165

Qy        219 KALLKRTAQAIEPYITTFFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLE 278
                                                   : :|  :|  :|::::   |:   :: |
Db        166 -------------------------RLNIEELLKHMYEVCQEMGLMEDLL--------- 190

Qy        279 FKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFAS 338
               ||    | |:   :|| |           :: |          |    :| |  |  :||
Db        191 -KLPFTDTEQEC-LVKFLQSSASVQNHEF------LLVHHLQRAN--YVP--------- 230

Qy        339 HCLMNHPDLAKDLTEYLKVR-SHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERT 397
                              |   |  : ||:     :|  :   :|                      :   |
Db        231 ---------ALKLNQTLKINVMNDRDPRLRE--------------------RSLARNSI 260

Qy        398 LDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVER 457
               ||:   ::          | ::::| |:: |                          |: ::      || |
Db        261 LDQYGKI-------LPRVHRKLAIERAKP----------------YHLSTSSVFRLVSR 296

Qy        458 IFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLR------------- 503
                        ||  :  |      :       ::           |  :   : |:|  :   :
Db        297 PKPLSAVPKQVVTGTVLTRSVFIN-----NVLSKIGEVWASKEPINSTTPFNSSKIEEPS 351

Qy        504 ----------------------HQVKDLLDLIKQPKTDASVKAIF-------SKVMVI 532
                                                      ::  ||||: ||     |   :  |       |  :  ::
Db        352 PIVYSLPAPELPEAFFGTPISKASQKISRLLDLVVQPVPRPSQCSEFIQQSSMKSPLYLV 411

Qy        533 TRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQ 592
               :|:||    : :     :    ::   |               | :|  :|         |:|:       :|
Db        412 SRSLPSSSQLKGSPQAISRASE--------LHLLETPLV-VKKAKSLAMSVT-------- 454

Qy        593 PTNPFLEMIKFLLERI---------------APVHIDTESISALIKQVN-KSIDGTADD 635
               |:  | |       :  |                   :|    :       || :  :  |:  |  |  |  |||
Db        455 -TSGFSEFTPQSILRSTPRSTPLASPSPSPGRSPQRLKETRISFVEEDVHPKWIPGAADD 513

Qy        636 EDEGV---PTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAALQIF 692
                   |     |     |:       |    |        |               :  :||:   :    |   :
Db        514 SKLEVFTTPKKCAVPVETEWPKSKDRTTSFFLNSPEKEHQ-----EMDEGSQSLEKLDVS 568

Qy        693 K-NTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPL 751
               | |:   |   |     :         |                   |     :    :         ||      :  |
Db        569 KGNSSVSITSDETTLEYQDAP---------SPEDLEETVFTASKPKSSSTALTTNVTEQT 619

Qy        752 HKSLDPSNLEHLITP---LVTIGHI-------ALLAPDQFA-----APWKSWVATFIVKD 796
                 |   |        :||         :|::          |:|  :  |:          :|   :  |:
Db        620 EKDGDKDVFASEVTPSDLQKQMGNLEDAETKDLLVAAEAFSELNHLSPVQGTEASLCAPS 679

Qy        797 LLMNDRLPGK-KTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKS-------G 847
               :           ||  |  |   ||     |   |  :  ::              :|            |  :||                 |
Db        680 V-----YEGKIFTQKSKVP---VLDEGLTSVETYTPAIR------ANDNKSMADVLGDGG 725

Qy        848 TSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQE--------PCYHEII 899
               |:|  :       : |:   |  ::   ::     :    :       | | :|                   |    ||
Db        726 NSSLTISEGPIVSERRLNQEVALNLKEDHEVEVGVLKESVDLPEEKLPISDSPPDTQEIH 785

Qy        900 TLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAI----CALCAKDPVKERR 955
               :|| :| |    |     :  |  :    :|:         |:|     :   |       | |          |
Db        786 VIEQEKLEA---QDSGEEARNLSFNELYPS-GTLKLQYNFDTIDQQFCDLADNKDTAE-- 839

Qy        956 AHARQCLVKNIN-------------------------VRREYLKQHA--AVSEKL---- 983
                     |  :   ::                                       : |  :  |   |   |||
Db        840 -----CDIAEVDGELFVAQSNFTLILEGEEGEVEPGDFASSDVLPKAANTATEEKLVCSG 894

Qy        984 -------LSLLPEYVV----PYTIHLLAHDPDYVKVQDIEQLKDV-------------- 1017
                         :: ||   |             :    |  :  |:  :||     | |  | ||
Db        895 ENDNHGQIANLPSAVTSDQKSQKVDTLPYVPEPIKVAIAENLLDVIKDTRSKEITSDTME 954

Qy       1018 ---KECLWFVLEILMAKNENNSHAF-----IRKMVENIKQTKDA--------------- 1053
                  |  :   |  :  :|       :      ||             |  :  |    |
Db        955 QSIHETIPLVSQNIMCPTKLVKSAFKTAQETSTMTMNVSQVDDVVSSKTRTRGQRIQNVN 1014

Qy       1054 ----------------QGPDDAKMNEKLYTVCDVAMNIIMS-KSTTYSLESPKDPVLPA 1095
                                 |      |       |      :  :  :  ||          :           :  :   |:: |  |
Db       1015 VKSAQQEASADVATPKMPGQSVRKKTRKAKEISEASENIYSDVRGLFQNQQIPQNSVTPR 1074

Qy       1096 RFFTQPDKNFSNTKNYLPPEMKSFFTPGKP----KTTNVL-GAVNK------PLSSAGKQ 1144
               |     :  :  |      :|        | :       |  :  |:   |::  :|    ||   :          :||   |:
Db       1075 RGRRKKEVNQDILENTSSVEQELQITTGRESKRLKSSQLLEPAVEETTKKEVKVSSVTKR 1134

Qy       1145 SQTKSSR----METVS--NASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKS 1198
               :   :   |         |:|      |         |    :||  |:    :|  |:   :|  |||                   |:
Db       1135 TPRRIKRSVENQESVEIINDLKVSTVTSPSRMIRKLRSTNLDASENTG--------NKQD 1186

Qy       1199 DKRDDSDLVRSELEKPRGRKKTPVTEQEE-KLGMDDLTKLVQEQKPKGSQRSRKRGHTAS 1257
               ||     |     |       :  :  |||:  :|         :|:          ||     ||        |      ||||
Db       1187 DKSSDKQLRIKHVRRVRGREVSPSDVREDSNLESSQLT--VQAEFDM-SAIPRKRGRPRK 1243

Qy       1258 ESDEQQWPEE----KRLKEDILENEDEQNSPPKKGKRGRPPK---------PLGGG--TP 1302
               :         |  |       |  :||:         :     |        |    ::    |    |    |                ||      |
Db       1244 IN-----PSEDVGSKAVKEERSPKKKEAPSIRRRSTRNTPAKSENVDVGKPALGKSILVP 1298

Qy       1303 KEEPTM---------KTSKKGSKKKSGPPAPEEEEEEERQSGNTEQKSK-------SKQH 1346
               || :|              | ::   |:|:|       ||    :|           ||: :                  :|
Db       1299 NEELSMVMSSKKKLTKKTESQSQKRSLHSVSEERTDEMTHKETNEQEERLLATASFTKSS 1358

Qy       1347 RVSRRAQQRA------ESPESSAIESTQS-TPQKGRGRPSKTPSPSQPKKNV 1391
               | ||       :|              |  :   :   |    |    |:|  :  :    |      |:|  |:  |
```

```
Db       1359 RSSRTRSSKAILLPDLSEPNNEPLFSPASEVPRKAKAK--KIEVPAQLKELV 1408

RESULT 6
US-10-118-513A-14
; Sequence 14, Application US/10118513A
; Publication No. US20030039995A1
; GENERAL INFORMATION:
;  APPLICANT: Taga, Tetsuya
;  APPLICANT:  Kimura, Naoki
;  TITLE OF INVENTION: THE YS68 GENE INVOLVED IN PRIMITIVE HEMATOPOIESIS
;  FILE REFERENCE: 06501-107US1
;  CURRENT APPLICATION NUMBER: US/10/118,513A
;  CURRENT FILING DATE:  2002-04-08
;  PRIOR APPLICATION NUMBER: PCT/JP00/05756
;  PRIOR FILING DATE: 2000-08-25
;  PRIOR APPLICATION NUMBER: JP 11-288738
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: JP 11-288739
;  PRIOR FILING DATE: 1999-10-08
;  PRIOR APPLICATION NUMBER: JP 2000-123721
;  PRIOR FILING DATE: 2000-04-19
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
;   LENGTH: 2266
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-118-513A-14

  Query Match             2.7%;  Score 197.5;  DB 15;  Length 2266;
  Best Local Similarity   18.2%;  Pred. No. 0.00078;
  Matches  293;  Conservative  223;  Mismatches  603;  Indels  493;  Gaps   68;

Qy         56 YLNLALHLASDFFL--------KHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIF 107
              |    :||  | :|         ||       :    | ||   : |    |  :    :|
Db        750 YPPASLHAVLDMYLLDGVTEAAKHS------ITIYLLLDIMYSF-PNKTDTPIESFPTVF 802

Qy        108 MFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINN 167
                    |:| ::            |:|:::  :     :: |     :   :  |   : : : : :
Db        803 AISWGQVKLIQG---------FWLIDHNDYESGLDLLFHPATAKPLSWQHSKIIQAFMSQ 853

Qy        168 GHNQKVHMHMVDLMSSIICEGDTVSQELLDTVLV---------NLVPAHKNLNKQAYDLA 218
              | ::: :  :   :  |   :   |:  |   |||:              |   :  |  |
Db        854 GEHRQA-LRYIQTMKPTVSSGNDVILHL--TVLLFNRCMVEAWNFLRQHCN--------- 901

Qy        219 KALLKRTAQAIEPYITTFFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLE 278
                                        : :| :| :|:::: |:  ::  |
Db        902 -------------------------RLNIEELLKHMYEVCQEMGLMEDLL--------- 926

Qy        279 FKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFAS 338
                ||   | |:   :|| |          :: |          |   :| | |  :||
Db        927 -KLPFTDTEQEC-LVKFLQSSASVQNHEF------LLVHHLQRAN--YVP---------- 966

Qy        339 HCLMNHPDLAKDLTEYLKVR-SHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERT 397
                       |  | : ||:    :| :  :|                          :  |
Db        967 ---------ALKLNQTLKINVMNDRDPRLRE--------------------RSLARNSI 996

Qy        398 LDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVER 457
              ||:  ::       | ::::| |:: |                     |: ::       || |
Db        997 LDQYGKI-------LPRVHRKLAIERAKP-----------------YHLSTSSVFRLVSR 1032

Qy        458 IFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLR-------------- 503
                      || :  |   :      :        |  :  |:|  :  :
Db       1033 PKPLSAVPKQVVTGTVLTRSVFIN-----NVLSKIGEVWASKEPINSTTPFNSSKIEEPS 1087

Qy        504 -----------------------HQVKDLLDLIKQPKTDASVKAIF-------SKVMVI 532
                                     ::  ||||: ||      |   :  |        | : ::
Db       1088 PIVYSLPAPELPEAFFGTPISKASQKISRLLDLVVQPVPRPSQCSEFIQQSSMKSPLYLV 1147

Qy        533 TRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQ 592
              :|:||    : :    :  ::   |             | : | :|     |:|:      :|
Db       1148 SRSLPSSSQLKGSPQAISRASE--------LHLLETPLV-VKKAKSLAMSVT-------- 1190

Qy        593 PTNPFLEMIKFLLERI---------------APVHIDTESISALIKQVN-KSIDGTADD 635
              |: | |        :  |                  :|  :       || :  : |: |  | | |||
Db       1191 -TSGFSEFTPQSILRSTPRSTPLASPSPSPGRSPQRLKETRISFVEEDVHPKWIPGAADD 1249

Qy        636 EDEGV---PTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAALQIF 692
                  |    |   |:   |   |    |         :| |        :||:   :     | :
Db       1250 SKLEVFTTPKKCAVPVETEWPKSKDRTTSFFLNSPEKEHQ-----EMDEGSQSLEKLDVS 1304

Qy        693 K-NTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPL 751
              | |:   |   |     :          |                |   :    :        ||        : |
Db       1305 KGNSSVSITSDETTLEYQDAP---------SPEDLEETVFTASKPKSSSTALTTNVTEQT 1355

Qy        752 HKSLDPSNLEHLITP---LVTIGHI-------ALLAPDQFA-----APWKSWVATFIVKD 796
               |     |      :||        :|::         |:| : |:        :| :    |:
Db       1356 EKDGDKDVFASEVTPSDLQKQMGNLEDAETKDLLVAAEAFSELNHLSPVQGTEASLCAPS 1415

Qy        797 LLMNDRLPGK-KTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKS--------G 847
              :              ||  |  |  ||    |   |  :  ::        :|             | :||              |
Db       1416 V-----YEGKIFTQKSKVP---VLDEGLTSVETYTPAIR------ANDNKSMADVLGDGG 1461

Qy        848 TSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQE--------PCYHEII 899
                |:| :      : |:   | ::  ::   :     :  :                | | :|         |   ||
Db       1462 NSSLTISEGPIVSERRLNQEVALNLKEDHEVEVGVLKESVDLPEEKLPISDSPPDTQEIH 1521

Qy        900 TLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAI----CALCAKDPVKERR 955
               :|| :| |         |     :  |  :     :|:        |:|    :    |        | |             |
Db       1522 VIEQEKLEA---QDSGEEARNLSFNELYPS-GTLKLQYNFDTIDQQFCDLADNKDTAE-- 1575

Qy        956 AHARQCLVKNIN------------------------VRREYLKQHA--AVSEKL---- 983
                    |   :   ::                                  :   |  :  |    |   |||
Db       1576 -----CDIAEVDGELFVAQSNFTLILEGEEGEVEPGDFASSDVLPKAANTATEEKLVCSG 1630

Qy        984 -------LSLLPEYVV----PYTIHLLAHDPDYVKVQDIEQLKDV-------------- 1017
                      :: ||   |           :   |  : |: :||   | ||
Db       1631 ENDNHGQIANLPSAVTSDQKSQKVDTLPYVPEPIKVAIAENLLDVIKDTRSKEITSDTME 1690

Qy       1018 ---KECLWFVLEILMAKNENNSHAF-----IRKMVENIKQTKDA--------------- 1053
                 |  :  | : :|    :       ||               |    | :  |    |
Db       1691 QSIHETIPLVSQNIMCPTKLVKSAFKTAQETSTMTMNVSQVDDVVSSKTRTRGQRIQNVN 1750

Qy       1054 -----------------QGPDDAKMNEKLYTVCDVAMNIIMS-KSTTYSLESPKDPVLPA 1095
                              |     |     |     :  :  :  ||        :        :  : |:: |  |
Db       1751 VKSAQQEASADVATPKMPGQSVRKKTRKAKEISEASENIYSDVRGLFQNQQIPQNSVTPR 1810

Qy       1096 RFFTQPDKNFSNTKNYLPPEMKSFFTPGKP----KTTNVL-GAVNK------PLSSAGKQ 1144
                  |   :  :  |         :|      |           |  :         |:: :|   || |:        :|| :
Db       1811 RGRKKEVNQDILENTSSVEQELQITTGRESKRLKSSQLLEPAVEETTKKEVKVSSVTKR 1870

Qy       1145 SQTKSSR----METVS--NASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKS 1198
                  :   :   |       |:|  |     |   :||  |:    :|  |:  :|  |||                  |:
Db       1871 TPRRIKRSVENQESVEIINDLKVSTVTSPSRMIRKLRSTNLDASENTG--------NKQD 1922

Qy       1199 DKRDDSDLVRSELEKPRGRKKTPVTEQEE-KLGMDDLTKLVQEQKPKGSQRSRKRGHTAS 1257
              ||  |   |       :   :   |||: :|     :|:    |        ||   ||  :        |   ||||
Db       1923 DKSSDKQLRIKHVRRVRGREVSPSDVREDSNLESSQLT--VQAEFDM-SAIPRKRGRPRK 1979

Qy       1258 ESDEQQWPEE----KRLKEDILENEDEQNSPPKKGKRGRPPK---------PLGGG--TP 1302
                  :     |  |       |  :||:      :  |     |   ::    |  |  |                ||      |
Db       1980 IN-----PSEDVGSKAVKEERSPKKKEAPSIRRRSTRNTPAKSENVDVGKPALGKSILVP 2034

Qy       1303 KEEPTM---------KTSKKGSKKKSGPPAPEEEEEEERQSGNTEQKSK-------SKQH 1346
                  || :|              |  ::     |:|:|            ||    :|                ||:  :       :|
Db       2035 NEELSMVMSSKKKLTKKTESQSQKRSLHSVSEERTDEMTHKETNEQEERLLATASFTKSS 2094

Qy       1347 RVSRRAQQRA------ESPESSAIESTQS-TPQKGRGRPSKTPSPSQPKKNV 1391
               | ||      :|    '       |   :   |    |  |:|  :   :    |        |:| |: |
Db       2095 RSSRTRSSKAILLPDLSEPNNEPLFSPASEVPRKAKAK--KIEVPAQLKELV 2144

RESULT 7
US-10-171-311-2
; Sequence 2, Application US/10171311
```

```
; Publication No. US20030087270A1
; GENERAL INFORMATION:
;  APPLICANT: Schlegel, Robert
;  APPLICANT:  Chen, Yan
;  APPLICANT:  Zhao, Xumei
;  APPLICANT:  Monahan, John
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Glatt, Karen
;  APPLICANT:  Gannavarapu, Manjula
;  APPLICANT:  Hoersh, Sebastian
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
;  TITLE OF INVENTION:  OF CERVICAL CANCER
;  FILE REFERENCE: MRI-035
;  CURRENT APPLICATION NUMBER: US/10/171,311
;  CURRENT FILING DATE:  2002-06-12
;  PRIOR APPLICATION NUMBER: US 60/298,159
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,155
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/335,936
;  PRIOR FILING DATE: 2001-11-14
;  NUMBER OF SEQ ID NOS: 238
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 3907
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-171-311-2

  Query Match             2.6%;  Score 189;  DB 15;  Length 3907;
  Best Local Similarity   18.5%;  Pred. No. 0.0063;
  Matches  283;  Conservative  267;  Mismatches  564;  Indels  416;  Gaps   73;

Qy         19 VKEISDK-----ISKEEMVRRLKMVVKTFMDMDQDSEEEKEL-----YLNLALHLASDFF 68
Db       2118 LKEKTDKCSELLLSKEQLQR---------DIQERNEEIEKLEFRVRELEQALLVSADTF 2167

Qy         69 LKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRY 128
Db       2168 QK------------VEDRKHFGAVEA---KPELSLEV------QLQAERDAIDRKEKEI 2205

Qy        129 FYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSSIICEG 188
Db       2206 TNLEEQLEQFRE-----ELENKNE---------------EVQQLHMQL----------- 2233

Qy        189 DTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTF--FNQVLMLGKT 246
Db       2234 -EIQKKESTTRLQELEQENKLFKDDMEKLGLAIKESDAMSTQDQHVLFGKFAQIIQEKEV 2292

Qy        247 SISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDN--EERLQVVKLLAKMFGA-- 302
Db       2293 EIDQLNEQVTKL----------------QQQLKITTDNKVIEEKNELIRDLETQIECLM 2335

Qy        303 KDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSHDP 362
Db       2336 SDQECVKRNR---EEEIEQLNEVIEKLQQELANIGQKTSMNAHSLSEE----------- 2380

Qy        363 EEAIRH--DVIVSIVTAAKKDILLVNDHLL---NFVRERTLD----------KRWRVRK 406
Db       2381 ADSLKHQLDVVIAEKLALEQQVETANEEMTFMKNVLKETNFKMNQLTQELFSLKRERESV 2440

Qy        407 EAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHI----YYQN---------SIDDR 452
Db       2441 EKIQSIPENSVNVAIDHLSKDKPELEVVLTEDALKSLENQTYFKSFEENGKGSIINLETR 2500

Qy        453 LLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQ-NLLRHQVKDLLD 511
Db       2501 LL-----------QLESTVSAK---DLELTQCYKQIKDMQEQGQFETEMLQKKIVNLQK 2545

Qy        512 LIKQPKTDASVKAI-------FSKVMVITRNL-PDPGKAQDFMKKFTQVLEDDEKIRKQL 563
Db       2546 IVEEKVAAALVSQIQLEAVQEYAKFCQDNQTISSEPERTN--IQNLNQLREDE--LGSDI 2601

Qy        564 EVLVSPTCSCKQAEGCVREITKKLGNPKQPT----NPFLEMIKFLLERIAPVHIDTESIS 619
Db       2602 SAL---TLRISELESQVVEMHTSLILEKEQVEIAEKNVLEKEKKLLE------------ 2645

Qy        620 ALIKQVNKSIDGT----ADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAET--FESL 673
Db       2646 -----LQKLLEGNEKKQREKEKKRSPQD------VEVLK----TTTELFHSNEESGFFNE 2690

Qy        674 LACLKMDDEKVA-EAALQIFKNTGSKIEED-------FPHIRSALLPVLHHKSKKGPPRQ 725
Db       2691 LEALRA--ESVATKAELASYKEKAEKLQEELLVKETNMTSLQKDLSQVRDHLA------E 2742

Qy        726 AKYAIHCIHAIFSS--KETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAA 783
Db       2743 AKEKLSILEKEDETEVQESKKACMFEPLPIKLSKSIASQ------TDGTLKISSSNQ--- 2793

Qy        784 PWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVS------PETMVKIQAIKMMVRWLL 837
Db       2794 -----TPQILVKNAGIQINLQSECSS------EEVTEIISQFTEKIEKMQ--ELHAAEIL 2840

Qy        838 GMKNNHSKSGTSTLR---LLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPC 894
Db       2841 DMESRHI-SETETLKREHYVAVQL-----LKEECGTLKAVIQCLRSKEGSSIPELAHSDA 2894

Qy        895 Y--HEIITLEQ--------------YQLCALAINDECYQVRQVFAQKLHKGLSRLRLPL 937
Db       2895 YQTREICSSDSGSDWGQGIYLTHSQGFDIASEGRGEESESATDSFPKKI-KGLLR-AVHN 2952

Qy        938 EYMAICALC-------------AKDP-VKERRAH-----ARQCLVKNINVRRE-----Y 972
Db       2953 EGMQVLSLTESPYSDGEDHSIQQVSEPWLEERKAYINTISSLKDLITKMQLQREAEVYDS 3012

Qy        973 LKQHAAVS----EKLLSLLPEYVVPYTIHLLAHDPDY-----------------VKVQD 1010
Db       3013 SQSHESFSDWRGELLLALQQVFLEERSVLLAAFRTELTALGTTDAVGLLNCLEQRIQEQG 3072

Qy       1011 IEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPD------DAKMNEK 1064
Db       3073 VE-YQAAMECLQKADRRSLLSEIQALHAQMNGRKITLKREQESEKPSQELLEYNIQQKQS 3131

Qy       1065 LYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLP---------P 1114
Db       3132 QMLEMQVELSSMKDRATELQEQLSSEKMVVAELKSE----LAQTKLELETTLKAQHKHLK 3187

Qy       1115 EMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRL 1174
Db       3188 ELEAFRLEVKDKTDEV-HLLNDTLASEQKKSRELQWALE--------------KEKAKL 3231

Qy       1175 D-SSEMDHSENEDYTMSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDD 1233
Db       3232 GRSEERDKEELEDLKFS------------------LESQKQRNLQLNLLLEQQKQLLNES 3273

Qy       1234 LTKLVQEQKPKGSQRSRKRGHTAS-----ESDEQQWPEEKRLKE--DILENEDEQNSPPK 1286
Db       3274 QQKIESQRMLYDAQLSEEQGRNLELQVLLES------EKVRIREMSSTLDRERELHAQLQ 3327

Qy       1287 KGK---RGRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAP--EEEEEEERQSGNT-EQK 1340
Db       3328 SSDGTGQSRPPLP-------SEDLLKELQKQLEEKHSRIVELLNETEKYKLDSLQTRQQM 3380

Qy       1341 SKSKQ-HRVSRRAQQRAESPESSAIESTQS 1369
Db       3381 EKDRQVHRKTLQTEQEANTEGQKKMHELQS 3410

RESULT 8
US-10-171-311-6
; Sequence 6, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
;  APPLICANT: Schlegel, Robert
;  APPLICANT:  Chen, Yan
```

```
; APPLICANT:  Zhao, Xumei
; APPLICANT:  Monahan, John
; APPLICANT:  Kamatkar, Shubhangi
; APPLICANT:  Glatt, Karen
; APPLICANT:  Gannavarapu, Manjula
; APPLICANT:  Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION:  OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE:  2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 3925
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-171-311-6

  Query Match             2.6%;  Score 189;  DB 15;  Length 3925;
  Best Local Similarity   18.5%;  Pred. No. 0.0063;
  Matches  283;  Conservative  267;  Mismatches  564;  Indels  416;  Gaps   73;

Qy         19 VKEISDK-----ISKEEMVRRLKMVVKTFMDMDQDSEEEKEL-----YLNLALHLASDFF 68
Db       2118 LKEKTDKCSELLLSKEQLQR---------DIQERNEEIEKLEFRVRELEQALLVSADTF 2167

Qy         69 LKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRY 128
Db       2168 QK------------VEDRKHFGAVEA---KPELSLEV------QLQAERDAIDRKEKEI 2205

Qy        129 FYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSSIICEG 188
Db       2206 TNLEEQLEQFRE-----ELENKNE----------------EVQQLHMQL----------- 2233

Qy        189 DTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTF--FNQVLMLGKT 246
Db       2234 -EIQKKESTTRLQELEQENKLFKDDMEKLGLAIKESDAMSTQDQHVLFGKFAQIIQEKEV 2292

Qy        247 SISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDN--EERLQVVKLLAKMFGA-- 302
Db       2293 EIDQLNEQVTKL----------------QQQLKITTDNKVIEEKNELIRDLETQIECLM 2335

Qy        303 KDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSHDP 362
Db       2336 SDQECVKRNR---EEEIEQLNEVIEKLQQELANIGQKTSMNAHSLSEE------------ 2380

Qy        363 EEAIRH--DVIVSIVTAAKKDILLVNDHLL---NFVRERTLD----------KRWRVRK 406
Db       2381 ADSLKHQLDVVIAEKLALEQQVETANEEMTFMKNVLKETNFKMNQLTQELFSLKRERESV 2440

Qy        407 EAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHI----YYQN----------SIDDR 452
Db       2441 EKIQSIPENSVNVAIDHLSKDKPELEVVLTEDALKSLENQTYFKSFEENGKGSIINLETR 2500

Qy        453 LLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQ-NLLRHQVKDLLD 511
Db       2501 LL-----------QLESTVSAK---DLELTQCYKQIKDMQEQGQFETEMLQKKIVNLQK 2545

Qy        512 LIKQPKTDASVKAI-------FSKVMVITRNL-PDPGKAQDFMKKFTQVLEDDEKIRKQL 563
Db       2546 IVEEKVAAALVSQIQLEAVQEYAKFCQDNQTISSEPERTN--IQNLNQLREDE--LGSDI 2601

Qy        564 EVLVSPTCSCKQAEGCVREITKKLGNPKQPT----NPFLEMIKFLLERIAPVHIDTESIS 619
Db       2602 SAL---TLRISELESQVVEMHTSLILEKEQVEIAEKNVLEKEKKLLE------------ 2645

Qy        620 ALIKQVNKSIDGT----ADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAET--FESL 673
Db       2646 -----LQKLLEGNEKKQREKEKKRSPQD------VEVLK----TTTELFHSNEESGFFNE 2690

Qy        674 LACLKMDDEKVA-EAALQIFKNTGSKIEED-------FPHIRSALLPVLHHKSKKGPPRQ 725
Db       2691 LEALRA--ESVATKAELASYKEKAEKLQEELLVKETNMTSLQKDLSQVRDHLA------E 2742

Qy        726 AKYAIHCIHAIFSS--KETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAA 783
Db       2743 AKEKLSILEKEDETEVQESKKACMFEPLPIKLSKSIASQ------TDGTLKISSSNQ--- 2793

Qy        784 PWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVS------PETMVKIQAIKMMVRWLL 837
Db       2794 -----TPQILVKNAGIQINLQSECSS------EEVTEIISQFTEKIEKMQ--ELHAAEIL 2840

Qy        838 GMKNNHSKSGTSTLR---LLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPC 894
Db       2841 DMESRHI-SETETLKREHYVAVQL-----LKEECGTLKAVIQCLRSKEGSSIPELAHSDA 2894

Qy        895 Y--HEIITLEQ--------------YQLCALAINDECYQVRQVFAQKLHKGLSRLRLPL 937
Db       2895 YQTREICSSDSGSDWGQGIYLTHSQGFDIASEGRGEESESATDSFPKKI-KGLLR-AVHN 2952

Qy        938 EYMAICALC-------------AKDP-VKERRAH-----ARQCLVKNINVRRE-----Y 972
Db       2953 EGMQVLSLTESPYSDGEDHSIQQVSEPWLEERKAYINTISSLKDLITKMQLQREAEVYDS 3012

Qy        973 LKQHAAVS----EKLLSLLPEYVVPYTIHLLAHDPDY-----------------VKVQD 1010
Db       3013 SQSHESFSDWRGELLLALQQVFLEERSVLLAAFRTELTALGTTDAVGLLNCLEQRIQEQG 3072

Qy       1011 IEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPD------DAKMNEK 1064
Db       3073 VE-YQAAMECLQKADRRSLLSEIQALHAQMNGRKITLKREQESEKPSQELLEYNIQQKQS 3131

Qy       1065 LYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLP---------P 1114
Db       3132 QMLEMQVELSSMKDRATELQEQLSSEKMVVAELKSE----LAQTKLELETTLKAQHKHLK 3187

Qy       1115 EMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRL 1174
Db       3188 ELEAFRLEVKDKTDEV-HLLNDTLASEQKKSRELQWALE--------------KEKAKL 3231

Qy       1175 D-SSEMDHSENEDYTMSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDD 1233
Db       3232 GRSEERDKEELEDLKFS-----------------LESQKQRNLQLNLLEQQKQLLNES 3273

Qy       1234 LTKLVQEQKPKGSQRSRKRGHTAS-----ESDEQQWPEEKRLKE--DILENEDEQNSPPK 1286
Db       3274 QQKIESQRMLYDAQLSEEQGRNLELQVLLES------EKVRIREMSSTLDRERELHAQLQ 3327

Qy       1287 KGK---RGRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAP--EEEEEEERQSGNT-EQK 1340
Db       3328 SSDGTGQSRPPLP-------SEDLLKELQKQLEEKHSRIVELLNETEKYKLDSLQTRQQM 3380

Qy       1341 SKSKQ-HRVSRRAQQRAESPESSAIESTQS 1369
Db       3381 EKDRQVHRKTLQTEQEANTEGQKKMHELQS 3410

RESULT 9
US-10-205-219-169
; Sequence 169, Application US/10205219
; Publication No. US20030138803A1
; GENERAL INFORMATION:
;  APPLICANT: Warner-Lambert Company
;  APPLICANT:  Lee, Kevin
;  APPLICANT:  Dixon, Alistair
;  APPLICANT:  Brooksbank, Robert
;  APPLICANT:  Pinnock, Robert
;  TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
```

```
;  FILE REFERENCE: WL-A-018200
;  CURRENT APPLICATION NUMBER: US/10/205,219
;  CURRENT FILING DATE:  2002-07-24
;  PRIOR APPLICATION NUMBER: GB 0118354.0
;  PRIOR FILING DATE: 2001-07-27
;  NUMBER OF SEQ ID NOS: 197
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 169
;   LENGTH: 2649
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Pemphigoid antigen
US-10-205-219-169

  Query Match             2.6%;  Score 186.5;  DB 12;  Length 2649;
  Best Local Similarity   18.8%;  Pred. No. 0.0052;
  Matches  257;  Conservative  232;  Mismatches  506;  Indels  369;  Gaps   69;

Qy        193 QELLDTVLVNLVPAHKNLNKQAYDLAK----ALLKRTAQAIEPYIT------TFFNQVLM 242
              ||: : :|:   ||    |  :||  |      : : :   | :| :|         |||
Db        433 QEIAEQLLLENHPA--RLTIEAYRAAMQTQWSWILQLCQCVEQHIKENTAYFEFFNDA-- 488

Qy        243 LGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMFGA 302
                 :|    :: |  |    |: |      :   |   |     :  || ||    :  |
Db        489 ---KEATDYLRNLKDAIQRKYSCDRSSSIHKLEDL-VQESMEEKEELLQYKSTIANLMGK 544

Qy        303 KDS--ELASQNKPLWQCYLGRFNDIHVPIRL---------------ECVKFASHCLMNHP 345
                :   :|   :|          | |          :||:              |||        | |:
Db        545 AKTIIQLKPRNS---DCPL----KTSIPIKAICDYRQIEITIYKDDECV------LANNS 591

Qy        346 DLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVR 405
                ||         || |      ||:     |  ::      |: : |  |        :      ||     |
Db        592 HRAK-----WKVISPTGNEAMVPSVCFTVPPPNKEAVDLANRIEQQYQNVLTL---WH-- 641

Qy        406 KEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKL--LHIYYQNSIDDRLLVERIFAQYM 463
              |: : :   :    : | :    :   |  :| || |   |    :::   |        |  :
Db        642 -ESHINMKSVVSWHYLINEIDRIRASNVASIKTMLPGEHQQVLSNLQSR------FEDF- 693

Qy        464 VPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVK 523
                  || ::  :     :::  |         : :: |   |: :    ::||    :: : :  ||
Db        694 ----LEDSQESQ----VFSGSD------ITQLEKEVNVCKQYYQELLKSAEREEQEESVY 739

Qy        524 AIF-SKVMVI-----------TRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSP-- 569
              :: :  |:|  |                 |     |:    |     :     ::    ||::|:|| |
Db        740 NLYISEVRNIRLRLENCEDRLIRQIRTPLERDDLHESVFRITE-QEKLKKELERLKDDLG 798

Qy        570 --TCSCKQ------AEGCVREITKKLGNPKQPTNPFLEMIKFLLERIAPVHI---DTESI 618
                 |  |::         |    |   :  :|         |   |     |      :::: |::    :|::
Db        799 TITNKCEEFFSQAAASSSVPTLRSELNVVLQNMNQVYSMSSTYIDKLKTVNLVLKNTQAA 858

Qy        619 SALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLK 678
                ||:|      :          |:|  |  |:                          |:|:: ||
Db        859 EALVKLYETKL-----CEEEAVIADK----------------------NNIENLISTLK 890

Qy        679 MDDEKVAEAALQIFKNTGSKI-------EEDFPHIRSALLPVLHHKSK----------- 719
                :| |   |:|      ::       :| |   :   |   || |
Db        891 QWRSEVDEKR-QVFHALEDELQKAKAISDEMFKTYKERDLDFDWHKEKADQLVERWQNVH 949

Qy        720 ---------KGPPRQAKYAIHCIHAI---FSSKETQFAQIFEPLHKSLDPSNLEHLITP 766
                         :|  :   ||        | :      ||        :||         | |  : |  |
Db        950 VQIDNRLRDLEGIGKSLKYYRDTYHPLDDWIQQVETTQRKIQEN-----QPENSKTLATQ 1004

Qy        767 L----VTIGHIALLAP--DQFAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSP 820
              |       :  :   |  :        |:          :  : ||   |||  :              |  :  |   :: ||
Db       1005 LNQQKMLVSEIEMKQSKMDECQKYAEQYSAT--VKDYELQ-----TMTYRAMVDSQQKSP 1057

Qy        821 ETMVKIQ-AIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLR 879
                   ::|  :  ::::  : ::   :       |: : |:|   :     ||   :    :: : :: |  :
Db       1058 VKRRRMQSSADLIIQEFMDLRTRY----TALVTLMTQYIKFAGDSLK--RLEEEEIKRCK 1111

Qy        880 LAAGSAIVKLAQEPCYHEII------TLEQYQLCA----------------LAINDECYQ 917
                    : ::   | :::          ||   :|                  : :|   :
Db       1112 --------ETSEHGAYSDLLQRQKATVLENSKLTGKISELERMVAELKKQKSRVEEELPK 1163

Qy        918 VRQVFAQKLHK--------GLSRLRLPLEYMAICALCAKD-------PVKERRAHARQC- 961
              ||:    :| |          | ::|   |              ||         |:|: |  |:
Db       1164 VREAAENELRKQQRNVEDISLQKIRAESE--------AKQYRRELETIVREKEAAERELE 1215

Qy        962 LVKNINVRREYLKQHAAVSEKLLSL---LPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVK 1018
              |: : :   |     : ||| | ||:      | |       :|       |     |::| :|:  | |::
Db       1216 RVRQLTIEAE--AKRAAVEENLLNFRNQLEENT--FTRRTL---EDHLKRKDL-SLNDLE 1267

Qy       1019 ECLWFVLEILMAKNENNSH------------AFIRKMVENIKQTKDAQGPD---DAKMNE 1063
              :     ::| |  |  :|                    || ::: |  || |: |   :        |: |
Db       1268 QQKNKLMEELRRKRDNEEELLKLIKQMEKDLAFQKQVAE--KQLKEKQKIELEARRKITE 1325

Qy       1064 KLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPG 1123
                ||  :  |:                       ||  |      ||          :  |     |   |
Db       1326 IQYTCRENAL-----------------PVCP---ITQATSCRAVTG--LQQE------HD 1357

Qy       1124 KPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSE 1183
                |  |    :    |::  |::|  ::::        :      ||        ||      |  ||     ::|  :
Db       1358 KQKAEELKQQVDE-LTAANRKAEQDMRELTYELNALQLEKTSS--EEKARLLKDKLDETN 1414

Qy       1184 NEDYTMSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMD--DLTKLVQEQ 1241
                |        :      |        :::|  ::     |:     |   ||:       |  :  |:     :    ||  |:  :
Db       1415 NTLRCLKLEL-----ERKDQAEKGYSQQLRELGRQLNQTTGKAEEAMQEASDLKKIKRNY 1469

Qy       1242 K--------PKGSQRSR----KRGHTASESDEQ-------QWPEEKRL--------KEDI 1274
              :                ||   :               |    |   :|  :  |          :  :||  |           |  |
Db       1470 QLELESLNHEKGKLQREVDRITRAHAVAEKNIQHLNSQIHSFRDEKELERLQICQRKSDH 1529

Qy       1275 LENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEERQS 1334
              |:  :  |::                                    |   ::   |   ::|::                 ||  |:  :
Db       1530 LKEQFEKS--------------------HEQLLQNIK--AEKENNDKIQRLNEELEKSN 1566

Qy       1335 GNTEQ-KSK----SKQHRVSRRAQQRAESPESSAIESTQSTPQK 1373
                  |   |   |        ::|:  ::   || ::    :   :    |:   |:
Db       1567 ECAEMLKQKVEELTRQNNETKLMMQRIQAESENIVLEKQTIQQR 1610


RESULT 10
US-10-205-823-419
; Sequence 419, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
;  APPLICANT: Schlegel, Robert
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Endege, Wilson O.
;  APPLICANT:  Gannavarapu, Manjula
;  APPLICANT:  Gorbatcheva, Bella
;  APPLICANT:  Hoersch, Sebastian
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Wonsey, Angela M.
;  APPLICANT:  Glatt, Karen
;  APPLICANT:  Zhao, Xumei
;  APPLICANT:  Anderson, Dustin
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
;  TITLE OF INVENTION:  METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF PROSTATE CANCER
;  FILE REFERENCE: MRI-044
;  CURRENT APPLICATION NUMBER: US/10/205,823
;  CURRENT FILING DATE:  2002-07-25
;  PRIOR APPLICATION NUMBER: 60/307,982
;  PRIOR FILING DATE: 2001-07-25
;  PRIOR APPLICATION NUMBER: 60/314,356
;  PRIOR FILING DATE: 2001-08-22
;  PRIOR APPLICATION NUMBER: 60/325,020
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: 60/341,746
;  PRIOR FILING DATE: 2001-12-12
;  PRIOR APPLICATION NUMBER: 60/362,158
;  PRIOR FILING DATE: 2002-03-05
;  NUMBER OF SEQ ID NOS: 455
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 419
;   LENGTH: 1979
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-205-823-419

  Query Match             2.6%;  Score 186;  DB 15;  Length 1979;
  Best Local Similarity   19.1%;  Pred. No. 0.0037;
  Matches  290;  Conservative  226;  Mismatches  567;  Indels  434;  Gaps   65;

Qy         16 PPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELY----LNLALHLASDFFLKH 71
              | ||:    || :: :   | : |      |: | |:      ||    || |
Db        362 PSAVKQSDTMTEKERILAQSASVEEVFRLQQALSDAENEIMRLSSLNQDNSLAED----- 416

Qy         72 PGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRYFYL 131
                :::| :            :   |    | :| ::: | :  :           |  : :
Db        417 ---NLKLKMR------IEVLEKEKSLLSQEK-EELQMSLLK-----------LNNEYEV 454

Qy        132 LENIAWVKSYNICFELED----------SNEIFTQLYRTLFSVIN--NGHNQKVHMHMV 178
              ::: |  :  ::  || |             |:  ::   || : |   :  ||:    ||:
Db        455 IKSTA-TRDISLDSELHDLRLNLEAKEQELNQSISE-KETLIAEIEELDRQNQEATKHMI 512

Qy        179 ---DLMSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKAL------------- 221
                 | :|   |||::  :|   :         |: |      |: | |
Db        513 LIKDQLSKQQNEGDSIISKLKQDLNDEKKRVHQ-LEDDKMDITKELDVQKEKLIQSEVAL 571

Qy        222 --LKRTAQAIEPYITTFFNQVLMLGKTSIS------DLSEHVFDLILELYNIDSHLLLSV 273
                |   | | :| :     :|:     ::::|      :| ||:      ||   | : |: |:
Db        572 NDLHLTKQKLEDKVENLVDQLNKSQESNVSIQKENLELKEHIRQNEEELSRIRNELMQSL 631

Qy        274 LPQLEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLEC 333
                         :||   |: | :           |||  |: |            :   :::|
Db        632 NQDSNSNFKDTLLKEREAEVRNLKQNL----SELEQLNENL--------KKVAFDVKMEN 679

Qy        334 VKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFV 393
               |     |                         | :|| :           :: |  |: |
Db        680 EKLVLAC----------------------EDVRHQL---------EECLAGNNQL---- 703

Qy        394 RERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRL 453
                 :|:|   |                      |:   |:  | :: | | :||        ::
Db        704 ---SLEKNTIVE--------------TLKMEKGEIEA-ELCWAKKRLL--------EEAN 737

Qy        454 LVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLI 513
              |:   :     || |:            :|    :|| | :             :: :|  |
Db        738 KYEKTIEELSNARNLNTSALQLEHEHL---IKLNQKKDM-----------EIAELKKNI 782

Qy        514 KQPKTDAS---------------VKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLEDDE 557
              :|  ||                    :  |  : |  : |    | |: : |::| |  :|
Db        783 EQMDTDHKETKDVLSSSLEEQKQLTQLINKKEIFIEKLKERSSKLQEELDKYSQALRKNE 842

Qy        558 KIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLLE---RIAPVHID 614
              :|: :|                  |  : ||: |:  |   | :: | |    | ||| |
Db        843 ILRQTIE-----------------EKDRSLGSMKEENNHLQEELERLREEQSRTAPV-AD 884

Qy        615 TESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLL 674
               ::: ::                             ||   :|  :  |  |  |  :
Db        885 PKTLDSV--------------------------TELASEVSQLNTIKEHLEEEIKH-- 914

Qy        675 ACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIH 734
                                       || ||    :  ||  |  : |:    : :|    |
Db        915 --------------------HQKIIEDQNQSKMQLLQSLQEQKKE--MDEFRYQ----H 947

Qy        735 AIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPWKSWVATFIV 794
                 ::  |   |:|  | |   :  :|:  | : |  |
Db        948 EQMNATHT---QLF--LEKDEEIKSLQKTIEQIKTQLH--------------------- 980

Qy        795 KDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKN---NHSKSGTSTL 851
                         ::    :  : ::  ||  |:|:          | ::|      |  |   |
Db        981 ----------EERQDIQTDNSDIFQET--KVQS--------LNIENGSEKHDLSKAETE 1019

Qy        852 RLLTTILHSDGD----------LTEQ-GKISKPDMSRLR-------LAAGSAIVKLAQEP 893
              ||:   |    : :            ||:|   ::||  :: :|        |    :     :::
Db       1020 RLVKGIKERELEIKLLNEKNISLTKQIDQLSKDEVGKLTQIIQQKDLEIQALHARISSTS 1079

Qy        894 CYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKE 953
               ::: |:| || |  |:  |     :|||    |      |     | ||  :  :  |    ||
Db       1080 HTQDVVYLQQ-QLQAYAMERE-----KVFAVLNEKTRENSHLKTEYHKMMDIVA---AKE 1130

Qy        954 RRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIE- 1012
                       |   | :: | |   |                    :   ||  |:    :: :|||
Db       1131 AALIKLQDENKKLSTRFESSGQD-------------MFRETIQNLSR---IIREKDIEI 1173

Qy       1013 -QLKDVKECLWFVLEILMAKNEN---NSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTV 1068
                |    : |  ||:        ||      |||  |       | :  | :|          || |     |
Db       1174 DALSQKCQTLLAVLQTSSTGNEAGGVNSHQF-----EELLQERDKLKQQVKKMEEWKQQV 1228

Qy       1069 CDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTT 1128
                  | :  :|      |  :    | |:       | |          :|      :     :   : |
Db       1229 MTTVQN-MQHESAQLQEELHQ---LQAQVLVDSDNNSKLQVDY--TGLIQSYEQNETKLK 1282

Qy       1129 NVLGAVNKPLSSAGKQSQTKS---SRMETVSNASSSSNPSSP--------GRIKGRLDSS 1177
                |      :  :      | |:    ||       ::: :|     ||:: :|                :  :  :||
Db       1283 NFGQELAQVQHSIGQLCNTKDLLLGKLDIISPQLSSASLLTPQSAECLRASKSEVLSESS 1342

Qy       1178 EMDHSENEDYTMSSPLPGKKSDKR---------DDSDLVRSELEKPRGRKKTPVTEQEEK 1228
              |:     | |:    |   |   | :  |             ||      ||||    ||:    |: | |
Db       1343 ELLQQELEELRKS--LQEKDATIRTLQENNHRLSDSIAATSELE----RKEHEQTDSEIK 1396

Qy       1229 L---GMDDLTKLVQEQ----KPKGSQRSRKRGHTASESDEQQWPEE--KRLKEDI----- 1274
                      | | ||::|:     | |  |             :  :: :| :    :       ||| |
Db       1397 QLKEKQDVLQKLLKEKDLLIKAKSDQLLSSNENFTNKVNENELLRQAVTNLKERILILEM 1456

Qy       1275 ----LENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAPEE---- 1326
                   |: |:|:        :||                      :|  || |     :|:       :  :|
Db       1457 DIGKLKGENEKIVETYRGKETE-------YQALQETNMKFSMMLREKEFECHSMKEKALA 1509

Qy       1327 -----EEEEERQSGNTEQ---KSKSKQHRVSRRAQQRAE---SPESSAIEST--QSTPQK 1373
                     :|:|:  ::|     |            ||| |  :       |:|:   :|     :    :|:|  |:  |:
Db       1510 FEQLLKEKEQGKTGELNQLLNAVKSMQEKTVVFQQERDQVMLALKQKQMENTALQNEVQR 1569

Qy       1374 GRGRPSKTPSPSQPKKN 1390
                |  :  ::      :  :|
Db       1570 LRDKEFRSNQELERLRN 1586

RESULT 11
US-09-815-242-5834
; Sequence 5834, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
```

```
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5834
;   LENGTH: 2437
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-5834

  Query Match             2.6%;  Score 185.5;  DB 9;  Length 2437;
  Best Local Similarity   18.5%;  Pred. No. 0.0054;
  Matches  237;  Conservative  187;  Mismatches  496;  Indels  361;  Gaps   47;

Qy        167 NGHNQKVHMHMVDLMSSIICEGDTVS--------------QELLDTVLVNLVP-AHKNL 210
              || ||:      ||    |    ||||               | |  :: | :| | : :
Db       1433 NGLNQQ----QQDLAHKAINNADTVSDVTDIVNNQIDLNDAMETLKHLVDNEIPNAEQTV 1488

Qy        211 NKQ-AYDLAKALL---KRTAQAIEPYITTFFNQVLMLGKTSISDLSEHVFDLILELYNID 266
              | | | | |  ||     || |           | :|    |:::|::  :   :  ::|:
Db       1489 NYQNADDNAKTNFDDAKRLA----------NTLLNSDNTNVNDINGAIQAVNDAIHNL- 1536

Qy        267 SHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIH 326
                               : ::|||         |||  : |  |    |    :  :|
Db       1537 -----------------NGDQRLQ---------DAKDKAIQSIN----QALANKLKEIE 1565

Qy        327 VPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTA----AKKDI 382
                                    | : |:  |: : :  | : : :| :|  |     |:  :
Db       1566 --------------------ASNATDQDKLIAKNKAEELANSIINNINKATSNQAESQV 1604

Qy        383 LLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLH 442
                 :| :  |     : |               |:|        | ||  ||:  : |::
Db       1605 QTAGNHAIEQVHANEIPK-----------AKI-------DANKDVDKQVQALIDEIDR 1644

Qy        443 IYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLL 502
                : : | :||      ||                          :|       |: : :  |
Db       1645 NPNLTDKEKQALKDRINQILQQGHN----------------GINNAMTKEEIEQAKAQL 1687

Qy        503 RHQVKDLLDLIK-----QPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVL---- 553
                ::|: ||:|       :     |  |:|:  ::       || |  | |      :  |:|
Db       1688 AQALQDIKDLVKAKEDAKQDVDKQVQALIDEIDQ-NPNLTDKEK-QALKDRINQILQQGH 1745

Qy        554 ---------EDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFL 604
                       |: |: : ||    :         :|:    :    | | | |:
Db       1746 NDINNALTKEEIEQAKAQLAQALQDIKDLVKAKEDAKNAIKALANAKR----------- 1793

Qy        605 LERIAPVHIDTESISALIKQVNKSIDGTADDEDEGV-PTDQAIRAGLELLKVLSFTHPIS 663
                                   |:| : | | : : : :   |:| :  |:            :
Db       1794 -----------------DQINSNPDLTPEQKAKALKEIDEAEKRALQ----------N 1824

Qy        664 FHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPP 723
                  :|:| : |    | :          |    :||      :: |
Db       1825 VENAQTIDQLNRGLNL--------GLDDIRNTHVWEVDEQP------------------ 1857

Qy        724 RQAKYAIHCIHAIFSSKETQFAQIFEPL-HKSLDPSNLEHLITPLVTIGHIALLAPDQFA 782
                       :: || :      |    | :|:      : :||      : || |:  || :
Db       1858 --------AVNEIFEATPEQILVNGELIVHR-------DDIITEQDILAHINLI--DQLS 1900

Qy        783 APWKSWVATFIVKDLLMN----DRLPGKKTTKLWVPDEEVSPE-TMVKIQAIKMMVRWLL 837
                |        :|   :  | |          | | |     : || :  |     | ::|| |||: :
Db       1901 AEVIDTPSTATISDSLTAKVEVTLLDGSKVI-VNVPVKVVEKELSVVKQQAIESIENAAQ 1959

Qy        838 GMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHE 897
                   |  : | | ||                  || : :  ::::|:     ||   :    |   |
Db       1960 QKINEINNSVTLTL---------------EQKEAAIAEVNKLK---QQAIDHVNDAPDVHS 2002

Qy        898 IITLEQYQLCAL-AINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRA 956
                : ::|  :   :  | | : |                            ||    :
Db       2003 VEEIQQQEQAHIEQFNPEQFTIEQ--------------------------AKSNAIKSIE 2036

Qy        957 HARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKD 1016
                | | ::  |  | :    :    :   |: | |    :                 :| : : :
Db       2037 DAIQHMIDEIKARTDLTDKEKQEAIAKLNQLKEQAIQ-------------AIQRAQSIDE 2083

Qy       1017 VKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNII 1076
                : | |      : | |          || | | : ||                |||: :: :  :
Db       2084 ISEQLEQFKAQMKAANPTAKELAKRKQ-EAISRIKDFS-------NEKINSIRNSEIGTA 2135

Qy       1077 MSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNK 1136
                |   ::     : ||     |  | | ::|      |       ::::     |  :    : ||
Db       2136 DEKQA--AMNQINEIVLE----TIRDINNAHTLQ----QVEAALNNGIAR----ISAVQI 2181

Qy       1137 PLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGK 1196
                |   |||         :   | |:|       |            :|| :  |           |
Db       2182 VTSDRAKQS-------SSTGNESNSHLTIGYGTANHPFNSSTIGHK------------K 2221

Qy       1197 KSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTA 1256
                | |: || | :           |              :   :|:   :: |:       ::| ||
Db       2222 KLDEDDDIDPLHM-----RHFSNNFGNVIKNAIGVVGISGLLASFWFFIAKRRRK----- 2271

Qy       1257 SESDEQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSK 1316
                | :|::        |: |  |  |:      |       : |   :       | :|: ::   :   |
Db       2272 -EDEEEELEIRDNNKDSIKETLDDTKHLPLLFVKRRRKEDEEDVTVEEKDSLNNGESLDK 2330

Qy       1317 KKSGP------PAPEEEEEEERQSGNTEQK-SKSKQHRVSRRAQQRAESPESSAIESTQS 1369
                |   |        |:||:  |   :  ||::|   |  :|      |::| :   :  :|:| |
Db       2331 VKHTPFFLPKRRRKEDEEDVEVTNENTDEKVLKDNEHSPLLFAKRRKDKEED--VETTTS 2388

Qy       1370 TPQKGRGRPSKTPSPSQPKKN 1390
                |     |            | |
Db       2389 IESKDEDVPLLLAKKKNQKDN 2409


RESULT 12
US-09-815-242-12996
; Sequence 12996, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12996
;   LENGTH: 6281
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12996

  Query Match             2.6%;  Score 185.5;  DB 9;  Length 6281;
  Best Local Similarity   18.5%;  Pred. No. 0.021;
  Matches  237;  Conservative  187;  Mismatches  496;  Indels  361;  Gaps   47;
```

```
Qy        167 NGHNQKVHMHMVDLMSSIICEGDTVS--------------QELLDTVLVNLVP-AHKNL 210
              || ||:         ||      |    ||||                 | |   :: | :| | : :
Db       5275 NGLNQQ----QQDLAHKAINNADTVSDVTDIVNNQIDLNDAMETLKHLVDNEIPNAEQTV 5330

Qy        211 NKQ-AYDLAKALL---KRTAQAIEPYITTFFNQVLMLGKTSISDLSEHVFDLILELYNID 266
              | | | | ||        || |                | :|   |:::|::  :  :   ::|:
Db       5331 NYQNADDNAKTNFDDAKRLA-----------NTLLNSDNTNVNDINGAIQAVNDAIHNL- 5378

Qy        267 SHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIH 326
                                 : ::|||          |||  : | |     |      :|
Db       5379 -----------------NGDQRLQ---------DAKDKAIQSIN----QALANKLKEIE 5407

Qy        327 VPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTA----AKKDI 382
                                        | : |:  |: : :  | : : :| :| |    |:  :
Db       5408 --------------------ASNATDQDKLIAKNKAEELANSIINNINKATSNQAESQV 5446

Qy        383 LLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLH 442
                    :| :  |     : |                |:|          | ||  ||:  : |::
Db       5447 QTAGNHAIEQVHANEIPK-----------AKI--------DANKDVDKQVQALIDEIDR 5486

Qy        443 IYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLL 502
                  :  :  | :||            ||                    :|      |: :  :  |
Db       5487 NPNLTDKEKQALKDRINQILQQGHN----------------GINNAMTKEEIEQAKAQL 5529

Qy        503 RHQVKDLLDLIK-----QPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVL---- 553
                ::|: ||:|       :    |  |:|:  ::        || |  | |      :  |:|
Db       5530 AQALQDIKDLVKAKEDAKQDVDKQVQALIDEIDQ-NPNLTDKEK-QALKDRINQILQQGH 5587

Qy        554 ---------EDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFL 604
                       |: |: : ||    :          :|:    :    | | | |:
Db       5588 NDINNALTKEEIEQAKAQLAQALQDIKDLVKAKEDAKNAIKALANAKR------------ 5635

Qy        605 LERIAPVHIDTESISALIKQVNKSIDGTADDEDEGV-PTDQAIRAGLELLKVLSFTHPIS 663
                                     |:|  : | |  : : : :   |:| :   |:         :
Db       5636 -----------------DQINSNPDLTPEQKAKALKEIDEAEKRALQ-----------N 5666

Qy        664 FHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPP 723
               :|:| : |     | :          |        :||           :: |
Db       5667 VENAQTIDQLNRGLNL--------GLDDIRNTHVWEVDEQP------------------ 5699

Qy        724 RQAKYAIHCIHAIFSSKETQFAQIFEPL-HKSLDPSNLEHLITPLVTIGHIALLAPDQFA 782
                            :: || :    |      | : |:          : :||     : || |:  || :
Db       5700 --------AVNEIFEATPEQILVNGELIVHR-------DDIITEQDILAHINLI--DQLS 5742

Qy        783 APWKSWVATFIVKDLLMN----DRLPGKKTTKLWVPDEEVSPE-TMVKIQAIKMMVRWLL 837
              |         :|  : | |              | | |     : || : |   | :|| |||: :
Db       5743 AEVIDTPSTATISDSLTAKVEVTLLDGSKVI-VNVPVKVVEKELSVVKQQAIESIENAAQ 5801

Qy        838 GMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHE 897
                 |     :  | | ||                 || : :   ::::|:      ||   :      |   |
Db       5802 QKINEINNSVTLTL-------------EQKEAAIAEVNKLK---QQAIDHVNDAPDVHS 5844

Qy        898 IITLEQYQLCAL-AINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRA 956
              :   ::|  :    |  | :  | :                              ||          :
Db       5845 VEEIQQQEQAHIEQFNPEQFTIEQ-------------------------AKSNAIKSIE 5878

Qy        957 HARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKD 1016
               | | ::  |  | :     :      :    |: | |  :                   :|   :  :  :
Db       5879 DAIQHMIDEIKARTDLTDKEKQEAIAKLNQLKEQAIQ-------------AIQRAQSIDE 5925

Qy       1017 VKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNII 1076
              : | |           : | |               || | | : ||              |||: :: :  :
Db       5926 ISEQLEQFKAQMKAANPTAKELAKRKQ-EAISRIKDFS-------NEKINSIRNSEIGTA 5977

Qy       1077 MSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNK 1136
                |     ::     : ||       |  | | ::|           ::::     |  :      : ||
Db       5978 DEKQA--AMNQINEIVLE----TIRDINNAHTLQ----QVEAALNNGIAR----ISAVQI 6023

Qy       1137 PLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGK 1196
                 :  ::|     :  | |:|        |            :|| : |                    |
Db       6024 VTSDRAKQS-------SSTGNESNSHLTIGYGTANHPFNSSTIGHK-------------K 6063

Qy       1197 KSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTA 1256
              | |: || | :       |              :   :|:  :: |:          ::| ||
Db       6064 KLDEDDDIDPLHM-----RHFSNNFGNVIKNAIGVVGISGLLASFWFFIAKRRRK----- 6113

Qy       1257 SESDEQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSK 1316
               | :|::          |: | |  |:    |      : |   :         | :|: ::   :   |
Db       6114 -EDEEEELEIRDNNKDSIKETLDDTKHLPLLFVKRRRKEDEEDVTVEEKDSLNNGESLDK 6172

Qy       1317 KKSGP------PAPEEEEEEERQSGNTEQK-SKSKQHRVSRRAQQRAESPESSAIESTQS 1369
              |    |          |:||: |  : ||::|   | :|        |::|  |   :|:| |
Db       6173 VKHTPFFLPKRRRKEDEEDVEVTNENTDEKVLKDNEHSPLLFAKRRKDKEED--VETTTS 6230

Qy       1370 TPQKGRGRPSKTPSPSQPKKN 1390
                  |    |              | |
Db       6231 IESKDEDVPLLLAKKKNQKDN 6251
```

RESULT 13
US-09-770-689A-5
; Sequence 5, Application US/09770689A
; Patent No. US20020115171A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001079
; CURRENT APPLICATION NUMBER: US/09/770,689A
; CURRENT FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 780
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-770-689A-5

```
  Query Match             2.6%;  Score 185;  DB 10;  Length 780;
  Best Local Similarity   18.8%;  Pred. No. 0.0011;
  Matches 154;  Conservative  131;  Mismatches  297;  Indels  236;  Gaps   37;

Qy        682 EKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCI---HAIFS 738
              | ::| ||        ||       :||    : || |:                | |     |    :
Db         23 EVLSEDLLQ--------IERRLDTVRS----MCHHSHKR--------LIACFQGQHGTDA 62

Qy        739 SKE------TQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPWKSWVATF 792
              :          |  ||  :       |: | |   :      | | |       | :| |     |       |
Db         63 ERRHKKLPLTALAQNMQEASAQLEESLLGKM---LETCGD----AENQLAFE-LSQHEVF 114

Qy        793 IVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRW---LLGMKNNHSKSGTS 849
              : |:::  | | |          : : |:        : |     :::: |             |   |||:
Db        115 VEKEIM--DPLYG-------IAEVEIPNIQKQRKQLARLVLDWDSVRARWNQAHKSSGTN 165

Qy        850 TLRLLTTILHSDGDLTEQGKISKPDMSRLRLAA-----------GSAIVKL--AQEPCY 895
                 | : |          :: | |  :|: |  : :|||                 |   |   |   ||    :
Db        166 FQGLPSKIDTLKEEMDEAG--NKVEQCKDQLAADMYNFMAKEGEYGKFFVTLLEAQADYH 223

Qy        896 HEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSR----LRLPLEYMAICALCAKDPV 951
               : : :  :   |   :  : : : :   :    |     | : | |       : ||:|   |      :
Db        224 RKALAVLEKALPEMRAHQDKWAEKPAFGTPLEEHLKRSGREIALPIE--ACVMLLLETGM 281

Qy        952 KER---RAHARQCLVKNINVRREYLKQH-----------AAVSEKLLSLLPEYVVPYTIH 997
               ||      |    |         :| :         :           |         |   :    |  ||| :: ::::
Db        282 KEEGLFRIGAGASKLKKLKAALDCSTSHLDEFYSDPHAVAGALKSYLRELPEPLMTFSLY 341

Qy        998 LLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPD 1057
                          :: :|  ::        : ||     :  |  :|   |    :: |  :  : ||  |
Db        342 -----EEWTQVASVQDQDKKLQYLWTTCQKLPPQNFVN-FRYLIKFLAKLAQTSDVN--- 392

Qy       1058 DAKMNEKLYTVCDVAM----NIIMSK----------STTYSLESPKDPVLP-ARFFTQPD 1102
                ||      |  ::|:       |:: :|                :|:   :  :  :|::  :|  :
Db        393 --KM-----TPSNIAIVLGPNLLWAKQEGTLAEIAAATSVHVVAVIEPIIQHADWFFPGE 445

Qy       1103 KNFSNTKNYLP-------------------------------PEMKSFFTPGKPK--- 1126
```

```
Db        446 VEFNVSEAFVPLATPNSNHSSHTGNDSDSGTLERKRPASMAVMEGDLVKKESPPKPKDSV 505

Qy       1127 --TTNVLGAVNKPLSSAGKQSQT-KSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSE 1183
Db        506 SAAAPVAGRNSNQITTVPNQAQTGGNSHQLSVGTAHSAAGPS-------------PHTL 551

Qy       1184 NEDYTMSSPLPGKKSD-----KRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLV 1238
Db        552 RRAVKKPAPAPPKPGNPPPGHPGGQSSPGTGTSPKPSTRSPSPPQQQQ----------- 599

Qy       1239 QEQKPKGSQRSRKRGHTASESDEQQWPEEKRLKEDILENEDEQNSP------PKKGKRGR 1292
Db        600 QQQQQQQQQQQQQQQQQQQQQQQQQTPGMRRCSSSLPPIQAPNHPPPQPPTQPRLGEQG- 658

Qy       1293 PPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEERQSGNTEQKSKSKQHRVSRRA 1352
Db        659 -PEP--GPTPPQTPT---------PPSTPPP---------------------------A 678

Qy       1353 QQRAESPESSAIESTQSTPQ---KGRGRPSKTPSPSQP 1387
Db        679 KQNSSQSETTQLHGTLPRPRPVPKPRNRPSVPPPPNPP 716
```

RESULT 14
US-10-128-714-3388
; Sequence 3388, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3388
;   LENGTH: 1814
;   TYPE: PRT
;   ORGANISM: Aspergillus fumigatus
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (1148)..(1148)
;   OTHER INFORMATION: Xaa = any amino acid
US-10-128-714-3388

```
  Query Match             2.6%;  Score 184.5;  DB 15;  Length 1814;
  Best Local Similarity   18.3%;  Pred. No. 0.0041;
  Matches  266;  Conservative  216;  Mismatches  546;  Indels  423;  Gaps   61;

Qy         48 DSEEEKELYLNLALHLASDFFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKD-- 105
Db        239 EKQKHEDIVISVLPILNDGFAMKNVSE---LIIGCYMVSV--VLAQKA--SLQDKVLDSL 291

Qy        106 -----------------IFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSYNICFELE 148
Db        292 MEAVAGSWTEETVESGLVCLAVLAQQK--PETKLPR-------------RALKAILRLD 335

Qy        149 DSNEIFTQL---YRT---LFSVINNGHNQKVHMHMVDLMSSIICEGDTVSQELLDTVLVN 202
Db        336 DIFKRLTELATQYKTSHLLLGVV-----------AGCLDDLPRQKDTARLDLLSQIFQN 383

Qy        203 LVPAHKNLNKQAYDLAKA--------LLKRTAQAIEPYITTFFNQVLMLGKTSISDLSEH 254
Db        384 QLLGEPEMSKAMALVLEATSSVHKDGAMSLDAQARLADLVQVFSQSESLRPTFQKTIAES 443

Qy        255 VFDLILELYNIDSHLLLSVLP------QLEFKLKSNDNEERLQVVKLLA--KMFGAKDSE 306
Db        444 SFDIAALEHNLQTVIDTAPAPRTVEDVEMEDVEKEEEQDHFSSTVESLSGEKLF--KGSF 501

Qy        307 LASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKD---LTEYLKVRSHDPE 363
Db        502 LSTQSIPLFEKLAQAFT-LAIGSKEKCQTFSDLAALGKNEAEKSPQYLSFFIRVFSGPYS 560

Qy        364 EAIR---HDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLA---QIYK 417
Db        561 MGTRVTALNMITSFLTSTESTNLDF-QALLPFLLVTLTDPSERVRREAAAALAAVGSLYK 619

Qy        418 KYALQSAAGKD---------AAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNL 468
Db        620 K----NRKGEDIWARDNLYGQPKDIKWLPTRDAQKVFERAVLPS-LEECIFDATHIGKVL 674

Qy        469 ETTER-------------MKCLYYLYA---------TLDLNAVKALNEMWKCQNLLR- 503
Db        675 ENTLRGVSVDANASELKKPLRLAFFTFLCSHAIELPLFTPKLGLLNILNRIDKAGGTTRT 734

Qy        504 --------------HQVKDLLDLIKQPKTDASVKAIF-------SKVMVITRNLPDP-- 539
Db        735 KELEPLLKTWRGFSEREVQDICEKERVPVSDVERQMVAIVTPKEKDSIMILLSNVSSHSE 794

Qy        540 ----------GKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLG 588
Db        795 SLRPSFIAAVFGRIKDI---WARVAEDRQTVAAEQLFDISLGLSDLPLVNNSRDLLRSVQ 851

Qy        589 NPKQPTNPFLEMIKFLLERIAPVHI-DTESISALIKQ----VNKSIDGTADDEDEGVPTD 643
Db        852 LPGSVLAQFLERI--------PVSLTDMEALGPAPKRRRTSQNNMIAMTVKDEAEFGKVM 903

Qy        644 QAIRAGLELLKVLS-FTHPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEED 702
Db        904 EKMTFILELVDSSSPETHP------ELADGLFQTL---------AALHHFK---SQIQSG 945

Qy        703 FPHIRS----ALLPVLHHKSKKGPPRQ-------AKYAIHCIHAIFSSKETQFAQIFEPL 751
Db        946 MSYLLSLTLGSLLAIV-NRSKESAKAQFDTSVIRADLVVDCVRTTESPQVQNAALLLVAG 1004

Qy        752 HKSLDPSNLEHLITPLVT-IGHIALLAPDQFAAPWKSWVATFIVKDLLMNDRLPGKKTTK 810
Db       1005 LSVIAPELVLHSVMPIFTFMGSSVLRKDDDYSVS------------------------- 1038

Qy        811 LWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKI 870
Db       1039 --VIDQTIDQVVPALIQSLRDQKRDVV--------SGTSELLLSFTAAFE--------- 1078

Qy        871 SKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQK----- 925
Db       1079 HIPSHRRLRL--------------FHALIT-------KLGTQDFLFAVLAMLANRYAMDK 1117

Qy        926 ----LHKGL-SRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVS 980
Db       1118 DVLVLMTGLVSDASAPVELTVRVVI----PRFWTXTDSKQTYSKFLGLVSDSLKPKPGIS 1173

Qy        981 EKLLSL------LPEYVVPYTI----HLLAHDPDYVKV------QDIEQLKDVKECLWFV 1024
Db       1174 QVLLGIGSDDGREPQKVAVDLLRDLAYLFKHSSLKVKMAKTFASEDEEAIRQLRSTFSQI 1233

Qy       1025 LEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYS 1084
Db       1234 LEQVLT---------IGDSVQSMKLVSQANGDVLAALFGTL-TLVDFLDTI------EVL 1277
```

```
Qy       1085 LESPKDP-------VLPARFFTQPDKNFSNTKNYLPPEMKSFFTP-------------- 1122
              || | |        :|  |    |::: ::     |       | |
Db       1278 LERPNDELRRKVLRLLEGRLRQNPERDSASQIRVLD------FLPTLVDIIRNSTDILLK 1331

Qy       1123 --------------GKPKTTNVLGAVNKPLSSA--GKQSQ--------TKSSRMETVSNA 1158
                            ||   :  |:||     | |  |:          :|  ||:
Db       1332 HAAVACIDRIAEKYGKKDPSRVIGAAQVVASEACIGQTDDRIRIMGVLCLASMAETLGQV 1391

Qy       1159 SSSSNPSSPGRIKGRLDSSEMDHSENE----------------DYTMSSP-------LP 1194
                 : | :  |     |: |  :  ||                 | :| |        |
Db       1392 MIPALPEALSRSLALLELSLEEGKENSRLHDAVFSLFSALFVHIPYMISGPHLDKILLLS 1451

Qy       1195 GKKSDKRDDSDLVRSELEKPRGRK------------------KTPVTEQEEKLGMDDLT 1235
                | ::  :  |  | |  |   ||                  ||  :|        ::
Db       1452 FKSANAEECEDDSRQEALKMMARKVDMAATLGAVDRNWQYAVQAGPVATKETL----EVV 1507

Qy       1236 KLVQEQKPKGS 1246
                |  |: || :
Db       1508 SLAVEKHPKSA 1518
```

RESULT 15
US-09-823-187-24
; Sequence 24, Application US/09823187
; Publication No. US20030096952A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Catherine
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Liu, Xiaohong
; APPLICANT: Majumder, Kumud
; APPLICANT: Padigaru, Muralidhar
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A
; APPLICANT: Spaderna, Steven K
; APPLICANT: Spytek, Kimberly
; APPLICANT: Taupier, Raymond J
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-745
; CURRENT APPLICATION NUMBER: US/09/823,187
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/193,339
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/193,205
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/195,343
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: 60/195,088
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,005
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,792
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: 60/196,556
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: 60/197,081
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/197,525
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/197,087
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-823-187-24

```
  Query Match             2.6%;  Score 184;  DB 11;  Length 676;
  Best Local Similarity   21.0%;  Pred. No. 0.0011;
  Matches   91;  Conservative   59;  Mismatches  146;  Indels  138;  Gaps   16;

Qy       1015 KDVKECLWFVLEILMAKNENNSHAF------IRKMVENIKQTKDAQGPDDAKMNEKLYTV 1068
              |   | || :          :|| ||         :       :    | | |  : :|
Db         75 KGFNEGLWEI--------QNNPHASYSAPPPVSSSDSEAPEANPADGSDADEDDEDRGV- 125

Qy       1069 CDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTT 1128
                :|:  : : : :  :||  |             ||:  |:             |  : |
Db        126 --MAVTAVTATAASDRMESDSD----------SDKSSDNS-------------GLKRKT 159

Qy       1129 NVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYT 1188
                |        |  |:::  || ::  |             ||    :     ||| : : ::|:|
Db        160 PALKV------SVSKRARKASSDLDQAS--------VSPSEEENSESSSESEKTSDQDFT 205

Qy       1189 ----------MSSPLPGKK---------SDKRDDSDLVRSE------------------ 1210
                          || |:|            || : |||  : |
Db        206 PEKKAAVRAPRRGPLGGRKKKKAPSASDSDSKADSDGAKPEPVAMARSASSSSSSSSSD 265

Qy       1211 ------------------LEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKR 1252
                                 | ||||||  |             |  :: :  : |    :|:|
Db        266 SDVSVKKPPRGRKPAEKPLPKPRGRKPKPERPPSSSSSDSDSDEVDRISEWKRRDEARRR 325

Qy       1253 GHTASESDEQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKE-----EPT 1307
                   |      ||:  | :||:|   : ::|:      ::  ||    :   || :  |       ||
Db        326 ELEARRRREQE-EELRRLRE---QEKEEKERRRERADRGEAERGSGGSSGDELREDDEPV 381

Qy       1308 MKTSKKGSKKKSGPPAPEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESPESSAIEST 1367
                |  :||   :   |||:  :  |  |                   :  |  |::  |:  |:||: |
Db        382 KKRGRKG--RGRGPPSSSDSEPEA---------------ELEREAKKSAKKPQSSSTEPA 424

Qy       1368 QSTPQK-GRGRPSK 1380
               :   ||   |  || :
Db        425 RKPGQKEKRVRPEE 438
```

Search completed: September 24, 2003, 20:30:40
Job time : 248 secs

```
OM protein - protein search, using sw model

Run on:         September 24, 2003, 20:16:09 ; Search time 25 Seconds
                                           (without alignments)
                                           2354.176 Million cell updates/sec

Title:          US-09-512-581B-2
Perfect score:  7193
Sequence:       1 MAHSKTRTNDGKITYPPGVK..........QKGRGRPSKTPSPSQPKKNV 1391

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters:      328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_AA:*
                1:  /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
                2:  /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
                3:  /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
                4:  /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
                5:  /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
                6:  /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	232.5	3.2	2662	4	US-09-595-684B-31	Sequence 31, Appl
2	201.5	2.8	3248	1	US-08-353-700-1	Sequence 1, Appli
3	201.5	2.8	3248	5	PCT-US95-16216-1	Sequence 1, Appli
4	194.5	2.7	2482	1	US-08-328-254-6	Sequence 6, Appli
5	190	2.6	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
6	185	2.6	8991	4	US-08-714-741-32	Sequence 32, Appl
7	181	2.5	1780	1	US-08-769-309A-5	Sequence 5, Appli
8	181	2.5	1780	3	US-08-994-570-5	Sequence 5, Appli
9	178.5	2.5	2368	1	US-08-198-446B-15	Sequence 15, Appl
10	178.5	2.5	2368	2	US-08-870-693-15	Sequence 15, Appl
11	176	2.4	1319	2	US-08-290-731C-2	Sequence 2, Appli
12	176	2.4	1336	2	US-08-290-731C-6	Sequence 6, Appli
13	174	2.4	1786	3	US-08-973-462-8	Sequence 8, Appli
14	172.5	2.4	1865	1	US-08-588-985-2	Sequence 2, Appli
15	172.5	2.4	1865	1	US-08-971-988-2	Sequence 2, Appli
16	170.5	2.4	3878	4	US-09-914-259-11	Sequence 11, Appl
17	170	2.4	1187	1	US-08-320-559-28	Sequence 28, Appl
18	170	2.4	1187	3	US-08-545-860D-28	Sequence 28, Appl
19	170	2.4	1187	5	PCT-US94-04496-28	Sequence 28, Appl
20	170	2.4	1210	1	US-08-320-559-26	Sequence 26, Appl
21	170	2.4	1210	3	US-08-545-860D-26	Sequence 26, Appl
22	170	2.4	1210	5	PCT-US94-04496-26	Sequence 26, Appl
23	168.5	2.3	3433	4	US-09-091-501B-10	Sequence 10, Appl
24	168	2.3	564	3	US-09-308-022-6	Sequence 6, Appli
25	166	2.3	947	4	US-09-418-780A-1	Sequence 1, Appli
26	166	2.3	1231	4	US-08-714-741-41	Sequence 41, Appl
27	165	2.3	1333	3	US-09-356-952-2	Sequence 2, Appli
28	165	2.3	1850	4	US-09-620-093A-5	Sequence 5, Appli
29	164.5	2.3	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
30	160	2.2	1312	4	US-09-345-882-29	Sequence 29, Appl
31	158.5	2.2	714	2	US-08-990-114-3	Sequence 3, Appli
32	158.5	2.2	714	4	US-09-241-333-3	Sequence 3, Appli
33	158.5	2.2	1507	4	US-09-914-259-37	Sequence 37, Appl
34	158	2.2	2474	4	US-08-265-967C-3	Sequence 3, Appli
35	158	2.2	2474	4	US-08-305-790B-4	Sequence 4, Appli
36	158	2.2	2972	4	US-09-579-181-2	Sequence 2, Appli
37	158	2.2	3118	4	US-09-579-181-1	Sequence 1, Appli
38	157.5	2.2	1111	4	US-09-914-259-28	Sequence 28, Appl
39	156.5	2.2	414	1	US-07-667-276A-4	Sequence 4, Appli
40	156.5	2.2	1588	5	PCT-US93-07261-11	Sequence 11, Appl
41	156.5	2.2	1663	5	PCT-US93-07261-16	Sequence 16, Appl
42	154.5	2.1	631	3	US-08-847-065-25	Sequence 25, Appl
43	154.5	2.1	1848	3	US-08-296-791-6	Sequence 6, Appli
44	154.5	2.1	1848	5	PCT-US95-10661A-6	Sequence 6, Appli
45	154.5	2.1	2101	1	US-08-466-390-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT 1
US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Ohashi, Cara
;  APPLICANT:  Sakowicz, Roman
;  APPLICANT:  VAisberg, Eugeni
;  APPLICANT:  Wood, Kenneth
;  APPLICANT:  Yu, Ming
;  TITLE OF INVENTION: Human kinesins and methods of producing
;  TITLE OF INVENTION:  and purifying human kinesins
;  FILE REFERENCE: cytop036
;  CURRENT APPLICATION NUMBER: US/09/595,684B
;  CURRENT FILING DATE:  2002-06-24
;  PRIOR APPLICATION NUMBER: 09/295,612
;  PRIOR FILING DATE: 2000-04-20
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
;   LENGTH: 2662
;   TYPE: PRT
;   ORGANISM: Human
US-09-595-684B-31

  Query Match              3.2%;  Score 232.5;  DB 4;  Length 2662;
  Best Local Similarity   18.2%;  Pred. No. 2.3e-09;
  Matches  274;  Conservative  270;  Mismatches  540;  Indels  425;  Gaps   64;

Qy         17 PGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEE----------EKELYLNLALHLASD 66
              | | |:|   : | :::| :   |  ||: :  ||          ||:    ||  |
Db        333 PYVNEVS---TDEALLKRYR---KEIMDLKKQLEEVSLETRAQAMEKD---QLAQLLEEK 383

Qy         67 FFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIF-MFITRQLKGL-EDTKSPQ 124
                |:                              :|::::  | :|     | :: |: :
Db        384 DLLQK-------------------------VQNEKIENLTRMLVTSSSLTLQQELKAKR 417

Qy        125 FNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSSI 184
                |  : |  |  :|: |               |   |::  |    | |   ::|:  |
Db        418 KRRVTWCLGKINKMKNSN---------------YADQFNIPTN-ITTKTHKLSINLLREI 461

Qy        185 ---ICEGDTVSQELLDTVL-VNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTFFNQV 240
                 :|    |    |||:  :   || | ||::  :
Db        462 DESVCSESDVFSNTLDTLSEIEWNPATKLLNQENIE------------------------ 497

Qy        241 LMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMF 300
                   : :: |     :|:|     |   | :   ::| |||  :: :  : ::   |
Db        498 -----SELNSLRADYDNLVL-----DYEQLRTEKEEMELKLKEKNDLDEFEALERKTK-- 545
```

```
Qy        301 GAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSH 360
              || |:                                |::    |:| :: :| :
Db        546 --KDQEMQ------------------------------LIHEISNLKNLVKHREVYNQ 571

Qy        361 DPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYA 420
              | | :           ::| ::|   :   :  ::|    ::     :    | |:      |:
Db        572 DLENEL---------SSKVELLREKEDQIKKLQEYIDSQK---LENIKMDLS-----YS 613

Qy        421 LQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYL 480
              |:|     ||  :  :  :              |  :  :  ::   |||  |:|| |
Db        614 LESIEDPKQMKQTLFDAETV-----------ALDAKRESAFLRSENLELKEKMKELATT 661

Qy        481 YATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITR------ 534
              |  ::      |::    |: |   :  |   :|| |:          ::: |::: :|
Db        662 YKQME-------NDIQLYQSQLEAKKKMQVDLEKE------LQSAFNEITKLTSLIDGKV 708

Qy        535 ------NLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLG 588
                    ||    ||  ||   |:   : :|::| :|:::  :|:|    |       |   : |::
Db        709 PKDLLCNLELEGKITDLQKELNKEVEENEALREEV-ILLS---ELKSLPSEVERLRKEI- 763

Qy        589 NPKQPTNPFLEMIKFLLERIAPVHIDTES-ISALIKQVNKSIDGTADDEDEGVPTDQAI- 646
                |  :  | :|       :::      :   ||  :   |:::: |: |  |   :      |||
Db        764 ---QDKSEELHIITSEKDKLFSEVVHKESRVQGLLEEIGKTKDDLATTQSNYKSTDQEFQ 820

Qy        647 --------------------RAGLELLKVLSFTHPISFHSAETFESLLACLKMD----- 680
                                  |    |:: :             |: |:|  |   || :
Db        821 NFKTLHMDFEQKYKMVLEENERMNQEIVNLSK--------EAQKFDSSLGALKTELSYKT 872

Qy        681 ---DEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKS-----------------K 719
                 || |   ::  |    :::|    :  |  |   |   |:                   :
Db        873 QELQEKTREVQERL--NEMEQLKEQLENRDSPLQTVEREKTLITEKLQQTLEEVKTLTQE 930

Qy        720 KGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPD 779
              |    :| : ::             ::      :  |  :  :: ::|     | |      |  :: |
Db        931 KDDLKQLQESLQI------ERDQLKSDIHDTVNMNIDTQ--EQLRNALESLKH------Q 976

Qy        780 QFAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGM 839
              :       || ::  : ::| |               || : ||  : |       : ::|:
Db        977 ETINTLKSKISEEVSRNLHM---------------EENTGETKDEFQ------QKMVGI 1014

Qy        840 KNNHSKSGTSTLRLLTTILHSDGDLTEQGKISK--PDMSRLRLAAGSAIVKLAQ------ 891
                         :| : ||   :   :  : :| ||        : : |:     | | :  |
Db       1015 DKKQDLEAKNT-QTLTADVKDNEIIEQQRKIFSLIQEKNELQQMLESVIAEKEQLKTDLK 1073

Qy        892 EPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCA--KD 949
              |       |    |: :|      : ||   :  ::: ||: :   :   :      |     |    |   ::
Db       1074 ENIEMTIENQEELRL----LGDELKKQQEIVAQEKNHAIKK---EGELSRTCDRLAEVEE 1126

Qy        950 PVKERRAHARQCLVKNINVRREYLKQHAAVSE--------KLLSLLPEYVVPYTIHLLAH 1001
              :||:    ::    :  :||: |     :      :|             |   |  |::     : |
Db       1127 KLKEKSQQLQEKQQQLLNVQEEMSEMQKKINEIENLKNELKNKELTLEHMETERLELAQK 1186

Qy       1002 -DPDYVKVQDI-EQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDA 1059
               : :| :|: | :: | :||        |:            ::      :||::      |||:
Db       1187 LNENYEEVKSITKERKVLKE-----LQKSFETERDHLRGYIREIEATGLQTKEELKIAHI 1241

Qy       1060 KMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSF 1119
               : |   |: ::  ::  |: |                   |:      |   |:||   | |:
Db       1242 HLKEHQETIDELRRSV--SEKT------------AQIINTQDLEKSHTK--LQEEIP-- 1282

Qy       1120 FTPGKPKTTNVLGAVNKPLSSAGKQSQTKS--SRMETVSNASSSSNPSSPGRIKGRLDSS 1177
                            ||     : | :   | |:|:    : :| ::   |:: : :: ||
Db       1283 ---------VLHEEQELLPNVKKVSETQETMNELELLTEQSTTKDSTTLARI------- 1325

Qy       1178 EMDHSE-NEDYTMSSPLPGKKSDKRDDSLVRSELEKPRGRKKTPVTE----------Q 1225
              ||:      || :   |     : :||:    ::   ||      :   |   :   |           |
Db       1326 EMERLRLNEKFQESQEEIKSLTKERDNLKTIKEALEVKHDQLKEHIRETLAKIQESQSKQ 1385

Qy       1226 EEKLGM----DDLTKLVQEQ---KPK-----------GSQRSRKRGH-----TASESDE 1261
              |: |  |         :: ||:| |       |||               |     :   :    |     | | |:
Db       1386 EQSLNMKEKDNETTKIVSEMEQFKPKDSALLRIEIEMLGLSKRLQESHDEMKSVAKEKDD 1445

Qy       1262 QQ------WPEEKRLKEDI-------LENEDEQNSPPKKGKRGRPP-KPLGGGTPKEEPT 1307
              |            |  :|||:|           || |:|          |           |         ::|
Db       1446 LQRLQEVLQSESDQLKENIKEIVAKHLETEEELKVAHCCLKEQEETINELRVNLSEKETE 1505

Qy       1308 MKTSKK-----GSKKKSGPPAPEEEEEE---ERQSGNTEQKSKSKQHRVSRRAQQRA-ES 1358
               :  | :|       | ::       |:||:    :: |     |   :: ||  :  |:|:   | :|
Db       1506 ISTIQKQLEAINDKLQNKIQEIYEKEEQLNIKQISEVQENVNELKQFKEHRKAKDSALQS 1565

Qy       1359 PESSAIEST 1367
               ||  :| |
Db       1566 IESKMLELT 1574
```

RESULT 2
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATTNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
US-08-353-700-1

Query Match 2.8%; Score 201.5; DB 1; Length 3248;
Best Local Similarity 19.5%; Pred. No. 1.2e-06;
Matches 298; Conservative 234; Mismatches 590; Indels 407; Gaps 63;

```
Qy         19 VKEISDKI--------SKEEMVRRLKMVVKTFMDMDQDSEEEKELY------------- 56
                :||:  |:          :| |     |:  :|        :  |  |: | :
Db       1900 MKELDSKLHLQEVQLMTKIEACIELEKIVGELKKENSDLSEKLEYFSCDHQELLQRVETS 1959

Qy         57 --LNLALHLASDFFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQL 114
                  ||  | : :|   |    :|:   ||                    |  |: |: :  :|
Db       1960 EGLNSDLEMHAD---KSSREDIGDNVA---------------KVNDSWKERFLDVENEL 2000

Qy        115 KGLEDTKSPQFNRYFYLLENIAWVKSYNICFELEDSN--EIFTQLYRTLFSVINNGHNQK 172
                  :    |:   :     ||   ::  |::   :| |  :: |   ::     |    | ||: :  ||
Db       2001 SRIRSEKASIEHEALYLEADLEVVQTEKLCLEKDNENKQKVIVCLEEEL-SVVTSERNQL 2059
```

```
Qy        173 VHMHMVDLMSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPY 232
                :| ||      | :|:::: :        : :|::          :|:| :|   : ::
Db       2060 --RGELDTMSKKTTALDQLSEKMKEK--TQELESHQSECLHCIQVAEAEVKEKTELLQ-- 2113

Qy        233 ITTFFNQVLMLGKTSISDLSEHVFDLI-----LELYNIDSHLLLSVLPQLEFKLKSNDNE 287
                          | ||:|| : |         |:    ||   |       :|| ::   : |
Db       2114 ------------TLSSDVSELLKDKTHLQEKLQSLEKDSQALSLTKCELENQIAQLNKE 2160

Qy        288 ERLQV--VKLLAKMFGAKDSELASQNKPLWQCYL--GRFNDIHVPIRLECVKFASHCLMN 343
              : | |     : |         | |  : :| |     :  | |        :||   :    | |
Db       2161 KELLVKESESLQARLSESDYEKLNVSKALEAALVEKGEF-----ALRLSSTQEEVHQLRR 2215

Qy        344 HPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKR-- 401
                          | |:||   |:   |   |    :   :::    : | :  |  ||  : :
Db       2216 G-------IEKLRVRIEADEKKQLH--IAEKLKERERENDSLKDKVENLERELQMSEENQ 2266

Qy        402 ---------------WRVRKEAMMGLAQIYKKYALQSAAGKD-AAKQIAWIKDKLLHIY 444
                              : : | |    :|::    :    : |:    |||   : :|  :
Db       2267 ELVILDAENSKAEVETLKTQIEEMARSLKIFELDLVTLRSEKENLTKQIQEKQGQLSELD 2326

Qy        445 YQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRH 504
                   |    || |:  |:   :     :|   |                            |::
Db       2327 KLLSSFKSLLEEKEQAEIQIKEESKTAVEM--------------------------LQN 2359

Qy        505 QVKDLLDLIKQPKTDASVKAIFSKVMVITRNLPDP-----GKAQDFMKKFTQVLEDDEKI 559
              |:|:| : :      |          ::|   |     ||          : :: ::|      || |||
Db       2360 QLKELNEAVAALCGD-------QEIMKATEQSLDPPIEEEHQLRNSIEKLRARLEADEK- 2411

Qy        560 RKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESIS 619
                 ||| ||       |::|      :   :: | ::       ||: :    |   |      :: |:
Db       2412 -KQLCVLQ----QLKESEHHADLLKGRVENLERE----LEIARTNQEHAA---LEAENSK 2459

Qy        620 ALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKM 679
                :: :    |:|                   |::| ||||          ::   | :   |:   |:
Db       2460 GEVETLKAKIEGMT----------QSLR-GLEL-------DVVTIRSEK--ENLTNELQK 2499

Qy        680 DDEKVAEAAL--QIFKNTGSKIEEDFPHIR---SALLPVLHHKSKKGPPRQAKYAIHCIH 734
              : |:::|   :      |:|   : |::    |   : :|  : |:     | |   |:|
Db       2500 EQERISELEIINSSFENILQEKEQEKVQMKEKSSTAMEMLQTQLKELNERVA--ALHNDQ 2557

Qy        735 AIFSSKETQFAQIFEPLH-------KSLDPSNLEHLITPLVTIGHIALLAPDQFAAPWKS 787
                 :||    :     | |           : || :     :::          |
Db       2558 EACKAKEQNLSSQVECLELEKAQLLQGLDEAKNNYIVLQSSVKG--------------- 2601

Qy        788 WVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHS--K 845
                             |: :    ||:  ||   |||:|        :||  : :|   |   ::   |   |
Db       2602 ---------LIQEVEDGKQ--KLEKKDEEIS-RLKNQIQDQEQLVSKLSQVEGEHQLWK 2648

Qy        846 SGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQ 905
                  || ||  |        :  :      |  |                            ||| |
Db       2649 EQNLELRNLTVELEQKIQVLQSKNASLQD-------------------------TLEVLQ 2683

Qy        906 LCALAINDECYQV---RQVFAQKLHKGLSR---LRLPLEYMAICALCAKDPV---KERRA 956
                      : :|               :  :  | :|::|   ::    |:  :  ||     :: :   | |
Db       2684 SSYKNLENELELTKMDKMSFVEKVNKMTAKETELQREMHEMAQKTAELQEELSGEKNRLA 2743

Qy        957 HARQCLVKNINVRREYLKQHAAVSEKLLSLLP-------------EYVVPYTIHLLAHD 1002
                  | |:: |     :: ||:     : :|     |                  | :   | :  |  |   |:
Db       2744 GELQLLLEEIKSSKDQLKELTLENSELKKSLDCMHKDQVEKEGKVREEIAEYQLRL--HE 2801

Qy       1003 PD------------YVKVQDI-EQLKDVKECL---WFVLEILMAKNE--NNSHAFIRK 1042
                 :                 |::|   |:|    :|||         :::|  :   |   |||         :
Db       2802 AEKKHQALLLDTNKQYEVEIQTYREKLTSKEECLSSQKLEIDLLKSSKEELNNSLKATTQ 2861

Qy       1043 MVENIKQTKDAQGPDDAKMNEKLYTVCDVA---MNIIMSKSTTYSLESPKDPV------L 1093
              ::| :|:||      |:   |       :|   :   |   | :::    :       ||    |:   :    |
Db       2862 ILEELKKTK----MDNLKYVNQLKKENERAQGKMKLLI--KSCKQLEEEKEILQKELSQL 2915

Qy       1094 PARFFTQPDKNFSNTK-NYLPPEMKSFFTPGKPKTTNV---------LGAVNKPLSSAGK 1143
                   |      |      :|| :  |  |:|          : ||                 :  |   |  :
Db       2916 QAAQEKQKTGTVMDTKVDELTTEIKELKETLEEKTKEADEYLDKYCSLLISHEKLEKAKE 2975

Qy       1144 QSQTK-----SSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLP---- 1194
                 :|:      | : :   |   |     |  ||                       ||:|
Db       2976 MLETQVAHLCSQQSKQDSRGSPLLGPVVPG--------------------PSPIPSVTE 3014

Qy       1195 -------GKKSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQ 1247
                            | | ||           ||           || |||                  |    | |:
Db       3015 KRLSSGQNKASGKRQ-----RSSGIWENGRGPTPAT---------------PESFSKKSK 3054

Qy       1248 RSRKRG-HTASESDEQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTP--KE 1304
              ::    | | |                       ||   | |    |:  |:|    | |   :|
Db       3055 KAVMSGIHPA---------------EDTEGTEFEPEGLPEVVKKGFADIPTGKTSPYILR 3099

Qy       1305 EPTMKT---SKKGSKKKSGPPAPEEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESPES 1361
                    ||| |      :  ::| :  |    :|       |   |     |:|:: :|::|          ||
Db       3100 RTTMATRTSPRLAAQKLALSPLSLGKENLAESSKPTAGGSRSQKVKVAQR------SPVD 3153

Qy       1362 SAI----ESTQSTPQKGRGRPSKTPSPSQ 1386
                   |         :|:| |          | |  || :
Db       3154 SGTILREPTTKSVPVNNLPERSPTDSPRE 3182
```

RESULT 3
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
;   APPLICANT: Yen, Timothy J.
;   APPLICANT: Rattner, Jerome B.
;   TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
;   TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Dann, Dorfman, Herrell and Skillman
;   STREET: 1601 Market Street Suite 720
;   CITY: Philadelphia
;   STATE: PA
;   COUNTRY: USA
;   ZIP: 19103-2307
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US95/16216
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/353,700
;   FILING DATE: 09-DEC-1995
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Reed, Janet E.
;   REGISTRATION NUMBER: 36,252
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (215) 563-4100
;   TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 3248 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: not relevant
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
PCT-US95-16216-1

```
  Query Match             2.8%;  Score 201.5;  DB 5;  Length 3248;
  Best Local Similarity   19.5%;  Pred. No. 1.2e-06;
  Matches  298;  Conservative  234;  Mismatches  590;  Indels  407;  Gaps   63;

Qy         19 VKEISDKI--------SKEEMVRRLKMVVKTFMDMDQDSEEEKELY------------- 56
               :||:  |:          :| |     |: :|       : |  |: | :
Db       1900 MKELDSKLHLQEVQLMTKIEACIELEKIVGELKKENSDLSEKLEYFSCDHQELLQRVETS 1959
```

```
Qy          57 --LNLALHLASDFFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQL 114
Db        1960 EGLNSDLEMHAD---KSSREDIGDNVA---------------KVNDSWKERFLDVENEL 2000

Qy         115 KGLEDTKSPQFNRYFYLLENIAWVKSYNICFELEDSN--EIFTQLYRTLFSVINNGHNQK 172
Db        2001 SRIRSEKASIEHEALYLEADLEVVQTEKLCLEKDNENKQKVIVCLEEEL-SVVTSERNQL 2059

Qy         173 VHMHMVDLMSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPY 232
Db        2060 --RGELDTMSKKTTALDQLSEKMKEK--TQELESHQSECLHCIQVAEAEVKEKTELLQ-- 2113

Qy         233 ITTFFNQVLMLGKTSISDLSEHVFDLI-----LELYNIDSHLLLSVLPQLEFKLKSNDNE 287
Db        2114 ------------TLSSDVSELLKDKTHLQEKLQSLEKDSQALSLTKCELENQIAQLNKE 2160

Qy         288 ERLQV--VKLLAKMFGAKDSELASQNKPLWQCYL--GRFNDIHVPIRLECVKFASHCLMN 343
Db        2161 KELLVKESESLQARLSESDYEKLNVSKALEAALVEKGEF-----ALRLSSTQEEVHQLRR 2215

Qy         344 HPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKR-- 401
Db        2216 G-------IEKLRVRIEADEKKQLH--IAEKLKERERENDSLKDKVENLERELQMSEENQ 2266

Qy         402 ---------------WRVRKEAMMGLAQIYKKYALQSAAGKD-AAKQIAWIKDKLLHIY 444
Db        2267 ELVILDAENSKAEVETLKTQIEEMARSLKIFELDLVTLRSEKENLTKQIQEKQGQLSELD 2326

Qy         445 YQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRH 504
Db        2327 KLLSSFKSLLEEKEQAEIQIKEESKTAVEM--------------------------LQN 2359

Qy         505 QVKDLLDLIKQPKTDASVKAIFSKVMVITRNLPDP-----GKAQDFMKKFTQVLEDDEKI 559
Db        2360 QLKELNEAVAALCGD-------QEIMKATEQSLDPPIEEEHQLRNSIEKLRARLEADEK- 2411

Qy         560 RKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESIS 619
Db        2412 -KQLCVLQ----QLKESEHHADLLKGRVENLERE----LEIARTNQEHAA---LEAENSK 2459

Qy         620 ALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKM 679
Db        2460 GEVETLKAKIEGMT----------QSLR-GLEL-------DVVTIRSEK--ENLTNELQK 2499

Qy         680 DDEKVAEAAL--QIFKNTGSKIEEDFPHIR---SALLPVLHHKSKKGPPRQAKYAIHCIH 734
Db        2500 EQERISELEIINSSFENILQEKEQEKVQMKEKSSTAMEMLQTQLKELNERVA--ALHNDQ 2557

Qy         735 AIFSSKETQFAQIFEPLH-------KSLDPSNLEHLITPLVTIGHIALLAPDQFAAPWKS 787
Db        2558 EACKAKEQNLSSQVECLELEKAQLLQGLDEAKNNYIVLQSSVKG--------------- 2601

Qy         788 WVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHS--K 845
Db        2602 ----------LIQEVEDGKQ--KLEKKDEEIS-RLKNQIQDQEQLVSKLSQVEGEHQLWK 2648

Qy         846 SGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQ 905
Db        2649 EQNLELRNLTVELEQKIQVLQSKNASLQD------------------------TLEVLQ 2683

Qy         906 LCALAINDECYQV---RQVFAQKLHKGLSR---LRLPLEYMAICALCAKDPV---KERRA 956
Db        2684 SSYKNLENELELTKMDKMSFVEKVNKMTAKETELQREMHEMAQKTAELQEELSGEKNRLA 2743

Qy         957 HARQCLVKNINVRREYLKQHAAVSEKLLSLLP-------------EYVVPYTIHLLAHD 1002
Db        2744 GELQLLLEEIKSSKDQLKELTLENSELKKSLDCMHKDQVEKEGKVREEIAEYQLRL--HE 2801

Qy        1003 PD-------------YVKVQDI-EQLKDVKECL---WFVLEILMAKNE--NNSHAFIRK 1042
Db        2802 AEKKHQALLLDTNKQYEVEIQTYREKLTSKEECLSSQKLEIDLLKSSKEELNNSLKATTQ 2861

Qy        1043 MVENIKQTKDAQGPDDAKMNEKLYTVCDVA---MNIIMSKSTTYSLESPKDPV------L 1093
Db        2862 ILEELKKTK----MDNLKYVNQLKKENERAQGKMKLLI--KSCKQLEEEKEILQKELSQL 2915

Qy        1094 PARFFTQPDKNFSNTK-NYLPPEMKSFFTPGKPKTTNV---------LGAVNKPLSSAGK 1143
Db        2916 QAAQEKQKTGTVMDTKVDELTTEIKELKETLEEKTKEADEYLDKYCSLLISHEKLEKAKE 2975

Qy        1144 QSQTK-----SSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLP---- 1194
Db        2976 MLETQVAHLCSQQSKQDSRGSPLLGPVVPG--------------------PSPIPSVTE 3014

Qy        1195 -------GKKSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQ 1247
Db        3015 KRLSSGQNKASGKRQ-----RSSGIWENGRGPTPAT--------------PESFSKKSK 3054

Qy        1248 RSRKRG-HTASESDEQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTP--KE 1304
Db        3055 KAVMSGIHPA--------------EDTEGTEFEPEGLPEVVKKGFADIPTGKTSPYILR 3099

Qy        1305 EPTMKT---SKKGSKKKSGPPAPEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESPES 1361
Db        3100 RTTMATRTSPRLAAQKLALSPLSLGKENLAESSKPTAGGSRSQKVKVAQR------SPVD 3153

Qy        1362 SAI----ESTQSTPQKGRGRPSKTPSPSQ 1386
Db        3154 SGTILREPTTKSVPVNNLPERSPTDSPRE 3182
```

```
RESULT 4
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
;  GENERAL INFORMATION:
;    APPLICANT:  Zhu, Xueliang
;    APPLICANT:  Lee, Wen-Hwa
;    TITLE OF INVENTION:  A No. 5710022el Nuclear Mitotic Phosphoprotein
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/328,254
;      FILING DATE:  24-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/141,239
;      FILING DATE:  22-OCT-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-CJ 1191
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2482 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-328-254-6

  Query Match             2.7%;  Score 194.5;  DB 1;  Length 2482;
```

```
Best Local Similarity   19.4%;  Pred. No. 2.8e-06;
Matches  297;  Conservative  233;  Mismatches  592;  Indels  407;  Gaps   63;

Qy        19 VKEISDKI--------SKEEMVRRLKMVVKTFMDMDQDSEEEKELY-------------- 56
Db      1172 MKELDSKLHLQEVQLMTKIEACIELEKIVGELKKENSDLSEKLEYFSCDHQELLQRVETS 1231

Qy        57 --LNLALHLASDFFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQL 114
Db      1232 EGLNSDLEMHAD---KSSREDIGDNVA---------------KVNDSWKERFLDVENEL 1272

Qy       115 KGLEDTKSPQFNRYFYLLENIAWVKSYNICFELEDSN--EIFTQLYRTLFSVINNGHNQK 172
Db      1273 SRIRSEKASIEHEALYLEADLEVVQTEKLCLEKDNENKQKVIVCLEEEL-SVVTSERNQL 1331

Qy       173 VHMHMVDLMSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPY 232
Db      1332 --RGELDTMSKKTTALDQLSEKMKEK--TQELESHQSECLHCIQVAEAEVKEKTELLQ-- 1385

Qy       233 ITTFFNQVLMLGKTSISDLSEHVFDLI-----LELYNIDSHLLLSVLPQLEFKLKSNDNE 287
Db      1386 -------------TLSSDVSELLKDKTHLQEKLQSLEKDSQALSLTKCELENQIAQLNKE 1432

Qy       288 ERLQV--VKLLAKMFGAKDSELASQNKPLWQCYL--GRFNDIHVPIRLECVKFASHCLMN 343
Db      1433 KELLVKESESLQARLSESDYEKLNVSKALEAALVEKGEF-----ALRLSSTQEEVHQLRR 1487

Qy       344 HPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDK--- 400
Db      1488 G-------IEKLRVRIEADEKKQLH--IAEKLKERERENDSLKDKVENLERELQMSEENQ 1538

Qy       401 -----------RWRVRKEAMMGLAQIYKKYALQ----SAAGKDAAKQIAWIKDKLLHIY 444
Db      1539 ELVILDAENSKAEVETLKTQIEEMARSLKVFELDLVTLRSEKENLTKQIQEKQGQLSELD 1598

Qy       445 YQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRH 504
Db      1599 KLLSSFKSLLEEKEQAEIQIKEESKTAVEM--------------------------LQN 1631

Qy       505 QVKDLLDLIKQPKTDASVKAIFSKVMVITRNLPDP-----GKAQDFMKKFTQVLEDDEKI 559
Db      1632 QLKELNEAVAALCGD-------QEIMKATEQSLDPPIEEEHQLRNSIEKLRARLEADEK- 1683

Qy       560 RKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESIS 619
Db      1684 -KQLCVLQ----QLKESEHHADLLKGRVENLERE----LEIARTNQEHAA---LEAENSK 1731

Qy       620 ALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKM 679
Db      1732 GEVETLKAKIEGMT---------QSLR-GLEL-------DVVTIRSEK--ENLTNELQK 1771

Qy       680 DDEKVAEAAL--QIFKNTGSKIEEDFPHIR---SALLPVLHHKSKKGPPRQAKYAIHCIH 734
Db      1772 EQERISELEIINSSFENILQEKEQEKVQMKEKSSTAMEMLQTQLKELNERVA--ALHNDQ 1829

Qy       735 AIFSSKETQFAQIFEPLH-------KSLDPSNLEHLITPLVTIGHIALLAPDQFAAPWKS 787
Db      1830 EACKAKEQNLSSQVECLELEKAQLLQGLDEAKNNYIVLQSSVNG--------------- 1873

Qy       788 WVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHS--K 845
Db      1874 ----------LIQEVEDGKQ--KLEKKDEEIS-RLKNQIQDQEQLVSKLSQVEGEHQLWK 1920

Qy       846 SGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQ 905
Db      1921 EQNLELRNLTVELEQKIQVLQSKNASLQD------------------------TLEVLQ 1955

Qy       906 LCALAINDECYQV---RQVFAQKLHKGLSR---LRLPLEYMAICALCAKDPV---KERRA 956
Db      1956 SSYKNLENELELTKMDKMSFVEKVNKMTAKETELQREMHEMAQKTAELQEELSGEKNRLA 2015

Qy       957 HARQCLVKNINVRREYLKQHAAVSEKLLSLLP-------------EYVVPYTIHLLAHD 1002
Db      2016 GELQLLLEEIKSSKDQLKELTLENSELKKSLDCMHKDQVEKEGKVREEIAEYQLRL--HE 2073

Qy      1003 PD-------------YVKVQDI-EQLKDVKECL---WFVLEILMAKNE--NNSHAFIRK 1042
Db      2074 AEKKHQALLLDTNKQYEVEIQTYREKLTSKEECLSSQKLEIDLLKSSKEELNNSLKATTQ 2133

Qy      1043 MVENIKQTKDAQGPDDAKMNEKLYTVCDVA---MNIIMSKSTTYSLESPKDPV------L 1093
Db      2134 ILEELKKTK----MDNLKYVNQLKKENERAQGKMKLLI--KSCKQLEEEKEILQKELSQL 2187

Qy      1094 PARFFTQPDKNFSNTK-NYLPPEMKSFFTPGKPKTTNV---------LGAVNKPLSSAGK 1143
Db      2188 QAAQEKQKTGTVMDTKVDELTTEIKELKETLEEKTKEADEYLDKYCSLLISHEKLEKAKE 2247

Qy      1144 QSQTK-----SSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLP---- 1194
Db      2248 MLETQVAHLCSQQSKQDSRGSPLLGPVVPG--------------------PSPIPSVTE 2286

Qy      1195 -------GKKSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQ 1247
Db      2287 KRLSSGQNKASGKRQ-----RSSGIWENGGGPTPAT--------------PESFSKKSK 2326

Qy      1248 RSRKRG-HTASESDEQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTP--KE 1304
Db      2327 KAVMSGIHPA---------------EDTEGTEFEPEGLPEVVKKGFADIPTGKTSPYILR 2371

Qy      1305 EPTMKT---SKKGSKKKSGPPAPEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESPES 1361
Db      2372 RTTMATRTSPRLAAQKLALSPLSLGKENLAESSKPTAGGSRSQKVKVAQR------SPVD 2425

Qy      1362 SAI----ESTQSTPQKGRGRPSKTPSPSQ 1386
Db      2426 SGTILREPTTKSVPVNNLPERSPTDSPRE 2454

RESULT 5
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
;   LENGTH: 10182
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

  Query Match             2.6%;  Score 190;  DB 4;  Length 10182;
  Best Local Similarity   19.1%;  Pred. No. 6.8e-05;
  Matches  276;  Conservative  235;  Mismatches  523;  Indels  414;  Gaps   67;

Qy        23 SDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLALHLASDFFLKHPGKDVRLLVAC 82
Db      8944 NDEIMNTNSIAQLSRIVNQAFDLNDAMKSLRDELNNQAFPVQASSNYINSDEDLKQQFDH 9003

Qy        83 CLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGL----EDTKSPQFNRYFYLLENIAWV 138
Db      9004 ALSNARKVLAKENGKNLDEK----------QIQGLKQVIEDTKD--------------- 9037

Qy       139 KSYNICFELEDSNEIFTQLYRTLFSVINNG--HNQKVHMHMVDLMSSI------ICEGDT 190
Db      9038 -ALNGIQRLSKAKAKAIQYVQSL-SYINDAQRHIAENNIHNSDDLSSLANTLSKASDLDN 9095
```

```
Qy        191 VSQELLDTV---------LVNLVPAHKNLN---KQAYDLAKALLKRT------------ 225
               ::| ||:            || : | |||      :|   | |    :|
Db       9096 AMKDLRDTIESNSTSVPNSVNYINADKNLQIEFDEALQQASATSSKTSENPATIEEVLGL 9155

Qy        226 AQAI---------EPYITTFFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQ 276
              :|||          |   :  |   :: |  |  |   : ||::    :  :     ::|   |: |  |
Db       9156 SQAIYDTKNALNGEQRLATEKSKDLKLIK-GLKDLNKAQLEDVTN--KVNSANTLTELSQ 9212

Qy        277 L-EFKLKSNDNEERLQ-VVKLLAKMFGAK------DSELASQ-NKPLWQC--YLGRFNDI 325
              | :  |: ||  : |:  :| |     |        |   |  | || | :      |  : :
Db       9213 LTQSTLELNDKMKLLRDKLKTLVNPVKASLNYRNADYNLKRQFNKALKEAKGVLNKNSGT 9272

Qy        326 HVPIRLECVKFASHCLMNHPDLAKD-LTEYLKVRSHDPEEAI----RHDVIVSIVTA--- 377
               :| |         | |:    | ||| |    ::: |   :    :       |:: :   |
Db       9273 NVNI-----NDIQH-LLTQIDNAKDQLNGERRLKEHQQKSEVFIIKELDILNNAQKAAII 9326

Qy        378 ----AKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYALQ-SAAGKDAAKQ 432
                   | ||| ::|   : | :                     :|| ||      |::  | :| ||
Db       9327 NQIRASKDIKIINQIVDNAIE---------LNDAMQGL----KEHVAQLTATTKD---- 9368

Qy        433 IAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLE---------TTERMKCLYYLYAT 483
                        :| | |: :| |        ||    ||            | :    :
Db       9369 ---------NIEYLNADEDHKL------QYDYAINLANNVLDKENGTNKDANIIGMIQN 9413

Qy        484 LD-----LNAVKALNE-MWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITRNLP 537
              :|      || :: |  :    |  | ::  :|  || |:        :: ||   :| ||
Db       9414 MDDARALLNGIERLKDAQTKAHNDIKDTLKRQLDEIEHANATSNSKA-QAKQMV------ 9466

Qy        538 DPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCS--CKQA--EG-------CVREITKK 586
                                :|: || |  :    | :      ||   ||             |: |
Db       9467 ------------------NEEARKALSNINDATSNDLVNQAKDEGQSAIEHIHADELPKA 9508

Qy        587 LGNPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAI 646
                 :    |    :| |   |:  :        :   :  || |:|| ::|   ::   :   :      |
Db       9509 KLDANQMIDQKVEDINHLISQ--NPNLSNEEKNKLISQINKLVNGIKNEIQQAINKQQIE 9566

Qy        647 RAGLELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHI 706
                |   :| :|:  |   :       ||                    | |: |    |  :
Db       9567 NATTKLDEVIETTKKLIIAKAE------------------AKQMIKELSQKKRD----- 9602

Qy        707 RSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDP--SNLEHLI 764
                       ::: :    | ::|     | :   |   :::        |          |:|     :| ||
Db       9603 ------AINNNTDLTPSQKA----HALADIDKTEKDALQHIEN--SNSIDDINNNKEHAF 9650

Qy        765 TPLVTIGHIALLAPDQ----FAAP----WKSWVATFIV------------KDLLMNDRL 803
                 |: || :    ||      |   |          : | : :|               | : : | |
Db       9651 N---TLAHIIIWDTDQQPLVFEVPELSLQNALVTSEVVVHRDETISLESIKKKMTLTDEL 9707

Qy        804 PGKKTTKLWVPD-EEVSPETMVKIQAI------------KMMVRWLLGMKNNHSKSGTS 849
                  |    : :|: ::|:          |:: |               |:: :  |     |  :    |:
Db       9708 ---KVNIVSLPNTDKVADHLTAKVKVILADGSYVTVNVPVKVVEKELQIAKKDAIKTIDV 9764

Qy        850 TLRLLTTILHSDGDLTE-QGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIIT-------- 900
               ::        : |: :||  |  : :| :: ||:        || |:        :| |
Db       9765 LVKQKIKDIDSNNELTSTQREDAKAEIERLK---KQAIDKVNHSKSIKDIETVKRTDFEE 9821

Qy        901 LEQYQLCALAIN--------DECYQVRQVFAQ-KLHKGLSRLRLPLEYMAICALCAKDPV 951
              ::|:       :|          |     |::   |  :  :    |||:            : || :
Db       9822 IDQFDPKRFTLNKAKKDIITDVNTQIQNGFKEIETIKGLTSNEKTQFDKQLTALQKEFLE 9881

Qy        952 KERRAHARQCLVKNINVRREY-------LKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPD 1004
               |  ||    ||:   :::|:           | |    : ||                      |:  |
Db       9882 KVEHAHN---LVELNQLQQEFNNRYKHILNQAHLLGEK-------------HIAEHKLG 9924

Qy       1005 YVKVQDIEQLKDVKECLWFVLE----------------ILMAKNENNSHAFIRKMVENIK 1048
              .|| |   :|: : :   :|: :                |    |      : |    : : :| |
Db       9925 YVVVNKTQQILNNQSASYFIKQWALDRIKQIQLETMNSIRGAHTVQDVHKALLQGIEQIL 9984

Qy       1049 QTKDA----------------QGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPV 1092
              :    :                          ||:: ||       ||| :|: :::      |:      |
Db       9985 KVNVSIINQSFNDSLHNFNYLHSKFDARLREK-----DVANHIVQTETFKEVLKG--TGV 10037

Qy       1093 LPARF--FTQPDKNFSNTKNYLPPEMKSFF--TPGKPKTTNVLGAVNKPLSSAGKQSQTK 1148
                  | :    ||  |   |  : |   :   |  | |     | :| :     |:   |: : :
Db      10038 EPGKINKETQQPKLHKNDNDSLFKHLVDNFGKTVGVITLTGLLSSFWLVLAKRRKKEEEE 10097

Qy       1149 SSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKRDDSDLVR 1208
                    ::                                                : ::|   :|        :|| |     : :
Db      10098 KQSIK---------------------------NHHKDIRLSD------TDKIDPIVITK 10123

Qy       1209 SELEKPRGRKKTPVTEQEEKLGMDDLTKL--VQEQKPKGSQRSRKRGHT------ASESD 1260
                  :::|                  ||::   ||      :     : :|  |  |  :   |:                  ||
Db      10124 RKIDK------------EEQIQNDDKHSIPVAKHKKSKEKQLSEEDIHSIPVVKRKQNSD 10171

Qy       1261 EQQWPEEK 1268
                  |::  |:
Db      10172 NQRYKTEE 10179
```

```
RESULT 6
US-08-714-741-32
; Sequence 32, Application US/08714741
; Patent No. 6500613
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E.
;    APPLICANT:  McDaniel, Larry S.
;    APPLICANT:  Swiatlo, Edwin
;    APPLICANT:  Yother, Janet
;    APPLICANT:  Crain, Marilyn J.
;    APPLICANT:  Hollingshead, Susan
;    APPLICANT:  Tart, Rebecca
;    APPLICANT:  Brooks-Walter, Alexis
;    TITLE OF INVENTION:  PNEUMOCOCCAL GENES, PORTIONS THEREOF,
;    TITLE OF INVENTION:  EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
;    TITLE OF INVENTION:  PORTIONS AND PRODUCTS
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford, P.C.
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/714,741
;      FILING DATE:  16-SEP-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer Esq., William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454312-2460
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8991 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  amino acid
US-08-714-741-32

  Query Match             2.6%;  Score 185;  DB 4;  Length 8991;
  Best Local Similarity   20.0%;  Pred. No. 0.00014;
  Matches  197;  Conservative  120;  Mismatches  322;  Indels  348;  Gaps   46;

Qy        538 DP-GKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQ-PTN 595
              || || || : |          : :: |::| | :        ::          |::|   | |   ||
Db       8210 DPEGKTQDELDKEAA----EAELNKKVEALPNQVAELEE------ELSKLEDNLKDAETN 8259
```

```
Qy    596 PFLEMIKFLLERIAPVHIDTESIS---ALIKQVNKSIDGT-----ADDEDEGVPTDQAIR 647
             : ||  ||          |:|:   | :::  | :|        | ::|  |  :|
Db   8260 HVEDYIKEGLE---------EAIATKQAELEKTPKELDAALNELGPDGDEEETPAPEA-- 8308

Qy    648 AGLELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAE----AALQIFKNTGSKIEEDF 703
                                   ||       :    || ||     |  :   |:  . :  |||:
Db   8309 ------------------PAE---------QPKPEKPAEETPAPAPKPEKSADQQAEEDY 8341

Qy    704 PHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHL 763
               ||      :  ::: ||:  | |           : |  | |    |   |      |||:|
Db   8342 AR-RSE--EEYNRLTQQQPPKAEKPA-----PAPAPKPEQPA----PAPKKKQKVNLENL 8389

Qy    764 ITPLVTIGHI------------------AL----------LAP--DQFAAPWKSWVATFI 793
           :: |    |                   ||           |:|   |          :  |   :|
Db   8390 LSTLDPGGKTQDELDKGAAEAELNKKVEALPNPVXELEEELSPPEDNLKDAETNHVEDYI 8449

Qy    794 VK--------------------DLLMNDRLP--GKKTTKLWV--PDEEV----------- 818
           :                      |  :|| :|  |:: |       |||
Db   8450 KEGLEEAIATKQAELEETPQEVDAALNDLVPDGGEEETPAPAPQPDEPAPAPAPNAEQPA 8509

Qy    819 ---SPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGD---LTEQ--GKI 870
              ||     || :   |                                  |:|:     ||:|    |
Db   8510 PAPKPEKSADQQAEEDYAR-----------------------RSEGEYNRLTQQQPPKA 8545

Qy    871 SKP----------------DMSRLRLAAGSAIVKLAQEPCYHEIIT--LEQYQLCALAI 911
           ||                   :::||:       : :|  |:|     : |      |||         ||
Db   8546 EKPAPAPAPKPEQPAPAPNKEIARLQ-----SDLKDAEENNVEDYIKEGLEQ------AI 8594

Qy    912 NDECYQV--RQVFAQKLHKGLSRLRLPLEYMAICALCAKDPV-KERRAHARQCLVKNINV 968
           :: ::   |      |  | |    | ||  :     |    ||   |  :      ::       :|
Db   8595 TNKKAELATTQQNIDKTQKDLEDAELELEKV----LATLDPEGKTQDELDKEAAEAELNE 8650

Qy    969 RREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEIL 1028
           : | |:    |  |: ||                          |::|    ||| :
Db   8651 KVEALQNQVAELEEELS--------------------KLED--NLKDAE---------- 8677

Qy   1029 MAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESP 1088
                  ||   :|:: :|      || |:              ||       |  |:|  :             :|
Db   8678 ----TNNVEDYIKEGLEEAIATKKAE-------LEKTQKELDAALNELGPDGDEEETPAP 8726

Qy   1089 KDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTK 1148
                   ||:|       :|  |                       ||   ||  :|::
Db   8727 A---------PQPEKPAEEPENPAP--------------------APKPEKSADQQAEED 8757

Qy   1149 SSRME-------------TVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYT------- 1188
           :|                    :  :   |   |       :::   |:|  ||:|||
Db   8758 YARRSEEEYNRLTQQQPPKAEKPAPAPQPEQPAPAP-KIELKEIDESESEDYAKEGFRAP 8816

Qy   1189 MSSPLPGKK---------SDKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQ 1239
           : |  |   ||               ||| |:  |                 :  :  |::|     :      |:
Db   8817 LHSKLDAKKAKLSKLEELSDKIDELD--------------AEIAKLEDQLKAVEENNNVE 8862

Qy   1240 EQKPKGSQRSRKRGHTASESDEQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGG 1299
           :   :| :::              |       |   |               ||: :  |   |         |              :|  ||
Db   8863 DYSTEGLEKTIAAKKTELEKTEAD------LKKAVNEPEKSAEEP------SQPEKPAEE 8910

Qy   1300 GTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESP 1359
              |:||   |   :    :::    | ||:   |:         :|:::     |       |         |
Db   8911 APAPEQPTEPTQPEKPAEETPAPKPEKPAEQPNAEKTDDQQAEEDYARRSEEEYNRL--- 8967

Qy   1360 ESSAIESTQSTPQKGRGRPSKTPSPSQ 1386
                    || |  |    :|:   |  |  |
Db   8968 -------TQQQPPKAE-KPAPAPQPEQ 8986
```

RESULT 7
US-08-769-309A-5
; Sequence 5, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauck, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/769,309A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5741890and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1780 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-769-309A-5

Query Match 2.5%; Score 181; DB 1; Length 1780;
Best Local Similarity 20.9%; Pred. No. 2.1e-05;
Matches 99; Conservative 59; Mismatches 151; Indels 164; Gaps 19;

```
Qy   1036 SHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPA 1095
           ||| |     |: :  :: :    : |                           |   : | ||| ||
Db    247 SHAEISPPAESGQAVEECKEEGEEK----------------QEKEPSKSAESPTSPVT-- 288

Qy   1096 RFFTQPDKNFSNTKNYLPPEMKSFFTPG-----------KPKTTNVLGAVNKPLSSAGKQ 1144
                     | | :       | ||| |                  |||   |   :   |         |
Db    289 ----------SETGS----TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKV 334

Qy   1145 SQTKSSRMETVS---NASSSSNPSSPG------RIKGRLDSSEMDHSENEDYTMS----- 1190
                 :  :| |      || :| |          |:    :   |:    |          :
Db    335 DTEEDGKAEVASEKLTASEQAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEK 394

Qy   1191 -SPLPGKKSDKRDD-------SDLVRSELEKPRGRKKTPVTE------QEEKLGMD---- 1232
           :||  :   |:: :          :::  | :|:    :|| |  |        || :|||
Db    395 PAPLATEVFDEKIEVHQEEVVAEVHVSTVEERTEEQKTEVEETAGSVPAEELVGMDAEPQ 454

Qy   1233 ------DLTKLVQ-----EQKPKGSQRSRKRGHTASESDEQQWPEEKRLKE--------- 1272
                 :| || :        |     :|: |                    |:|| |  :
Db    455 EAEPAKELVKLKETCVSGEDPTQGADLS----------------PDEKVLSKPPEGVVSEV 499

Qy   1273 DILENEDE---QNSPPK-----------------KGKRGRPPKPLGGGT--PKEEPTMKTS 1311
           ::| :::        | || |                     |||||   :   |   |   |  : |   :
Db    500 EMLSSQERMKVQGSPLKKLFTSTGLKKLSGKKQKGKRGGDEESGEHTQVPADSPDSQEE 559

Qy   1312 KKGSKKKSGPPAPEE-----------EEEEERQSGNTEQKSKSKQ--------------- 1345
           :||    |  |  |||                    ::: |  :|  |   ::
Db    560 QKGESSASSPEEPEEITCLEKGLAEVQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPK 619

Qy   1346 HRVSRRAQ-------QRAESPESSAIESTQSTPQKGRGRPSKTPSPSQPKKNV 1391
            || | ::               : :|     |:  ||| |  |:           : |  |  :||: |
Db    620 KRVRRPSESDKEDELDKVKSATLSSTESTASEMQEEMKGSVEEPKPEEPKRKV 672
```

RESULT 8
US-08-994-570-5

```
; Sequence 5, Application US/08994570
; Patent No. 6090929
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, John D.,
;    APPLICANT:  Nauert, Brian J.,
;    APPLICANT:  Klauck, Theresa M.
;    TITLE OF INVENTION:  Protein Binding Domains of Gravin
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower/233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/994,570
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090929and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/33451
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1780 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-994-570-5

  Query Match             2.5%;  Score 181;  DB 3;  Length 1780;
  Best Local Similarity   20.9%;  Pred. No. 2.1e-05;
  Matches   99;  Conservative   59;  Mismatches  151;  Indels  164;  Gaps   19;

Qy       1036 SHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPA 1095
              |||  |     |: :  :: :   : |                 |  : | |||  ||
Db        247 SHAEISPPAESGQAVEECKEEGEEK----------------QEKEPSKSAESPTSPVT-- 288

Qy       1096 RFFTQPDKNFSNTKNYLPPEMKSFFTPG-----------KPKTTNVLGAVNKPLSSAGKQ 1144
                        | | :       | ||| |            |||    |  :  |      |
Db        289 ----------SETGS----TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKV 334

Qy       1145 SQTKSSRMETVS---NASSSSNPSSPG------RIKGRLDSSEMDHSENEDYTMS----- 1190
                   :  : |  |      ||  ::|  |          |:    :   |:    |    :
Db        335 DTEEDGKAEVASEKLTASEQAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEK 394

Qy       1191 -SPLPGKKSDKRDD-------SDLVRSELEKPRGRKKTPVTE------QEEKLGMD---- 1232
               :||  :  |:: :          ::: | :|:      :|| |  |         || :|||
Db        395 PAPLATEVFDEKIEVHQEEVVAEVHVSTVEERTEEQKTEVEETAGSVPAEELVGMDAEPQ 454

Qy       1233 ------DLTKLVQ-----EQKPKGSQRSRKRGHTASESDEQQWPEEKRLKE--------- 1272
                    :| || :        |     :|: |                 |:|| | :
Db        455 EAEPAKELVKLKETCVSGEDPTQGADLS---------------PDEKVLSKPPEGVVSEV 499

Qy       1273 DILENEDE---QNSPPK----------------KGKRGRPPKPLGGGT--PKEEPTMKTS 1311
              ::| :::      | || |                |||||    :   |   |    |  : |   :
Db        500 EMLSSQERMKVQGSPLKKLFTSTGLKKLSGKKQKGKRGGGDEESGEHTQVPADSPDSQEE 559

Qy       1312 KKGSKKKSGPPAPEE-----------EEEEERQSGNTEQKSKSKQ--------------- 1345
               :||    | |  |||              ::: |  :  | |      | ::
Db        560 QKGESSASSPEEPEEITCLEKGLAEVQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPK 619

Qy       1346 HRVSRRAQ-------QRAESPESSAIESTQSTPQKGRGRPSKTPSPSQPKKNV 1391
               || | ::            : :|    |: ||| |  |:       : | | :||: |
Db        620 KRVRRPSESDKEDELDKVKSATLSSTESTASEMQEEMKGSVEEPKPEEPKRKV 672

RESULT 9
US-08-198-446B-15
; Sequence 15, Application US/08198446B
; Patent No. 5674996
;  GENERAL INFORMATION:
;    APPLICANT:  Hartwell, Leland H.
;    APPLICANT:  Weinert, Ted A.
;    APPLICANT:  Plon, Sharon E.
;    APPLICANT:  Groudine, Mark T.
;    TITLE OF INVENTION:  Cell Cycle Checkpoint Genes
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson Kindness PLLC
;      STREET:  1420 Fifth Ave., Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/198,446B
;      FILING DATE:  18-FEB-1994
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  FHCR17537
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-682-8100
;      TELEFAX:  206-224-0779
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2368 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;      DESCRIPTION:  yeast MEC1 protein
US-08-198-446B-15

  Query Match             2.5%;  Score 178.5;  DB 1;  Length 2368;
  Best Local Similarity   19.4%;  Pred. No. 5.5e-05;
  Matches  251;  Conservative  188;  Mismatches  477;  Indels  377;  Gaps   60;

Qy         68 FLKHPGKD--VRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQF 125
              || ||  |       |::  |   : ||::     :   | || |       |  :  |  |
Db        526 FLSHPNIDEFSESLLSGILFSLHRIFS----HFQPPKLTDGNGQINKSFK-LVQKCFMNS 580

Qy        126 NRYFYLLENIAWVKSYNI--CFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSS 183
              |||  ||         |   |    | : ||:      :    ||    :  :       |  ::|    :
Db        581 NRYLRLL-------STRIIPLFNISDSHNSEDEHTATLIKFLQSQKLPVVKENLVIAWTQ 633

Qy        184 IICEGDTVSQELLDTVLVNLVPAHKN--------LNKQAYDLAKALLKRTAQAIEPYITT 235
              :      |  | ::  ||:|:  |:         :          :   |  :: ||  |     |  :  :
Db        634 LTL---TTSNDVFDTLLLKLIDIFNSDDYSLRIMMTLQIKNMAKILKKTPYQLLSPILPV 690

Qy        236 FFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLLLS-----VLPQLEFKLKSNDNEERL 290
                  |    |||   :      :    |    ::||       |    :|         ::|      :  ||:
Db        691 LLRQ---LGKNLVE--RKVGFQNLIELLGYPSKTILDIFQRYIIPYAIIQYKSD------ 739

Qy        291 QVVKLLAKMFGAKDSELASQNKP-------------------------LWQCYLGR---F 322
               |:  :||:     |:  |  :|  |                                    |     :|  |     |
Db        740 -VLSEIAKIMCDGDTSLINQMKVNLLKKNSRQIFAVALVKHGLFSLDILETLFLNRAPTF 798

Qy        323 NDIHVPIRLECVK-FASHCLMNHPDLAKD-----------------LTEYLKVRSH---- 360
               :   ::    |    |    |       :       :  ||                       :|  :  |  :  |
```

```
Db        799 DKGYITAYLPDYKTLAEITKLYKNSVTKDASDSENANMILCSLRFLITNFEKDKRHGSKY 858

Qy        361 --------DPEEAIR---HDVIVSIVTAAKKDILLVNDHLLNFVRERTLDK-----RWRV 404
                      | |:| :      | |: |      || |        : : | ::         :
Db        859 KNINNWTDDQEQAFQKKLQDNILGIFQVFSSDIHDVEGRTTYEKLRVINGISFLIIYAP 918

Qy        405 RKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRL--LVERIFA-- 460
              :| : ||||     ||: |     |:: :   :  |:  ::  |: |   ::: : |
Db        919 KKSIISALAQI--SICLQTGLG---LKEVRYEAFRCWHLLVRHLNDEELSTVIDSLIAFI 973

Qy        461 -----------QYMVPHNLETTERMKCLYYL----YATLDLNAVKALNEMWKCQNLLR-- 503
                          : :|    |:|  : |    |     | || |      | : :      |
Db        974 LQKWSEFNGKLRNIVYSILDTLIKEKSDLILKLKPYTTLALVGKPELGILARDGQFARMV 1033

Qy        504 HQVKDLLDLIKQPKTDASVKAIFSKVMVITRNLPD-------------------PGKAQ 543
              ::::    |||   |       ::|:   |     || :|| |                 |:
Db       1034 NKIRSTTDLI--PIFANNLKS--SNKYVINQNLDDIEVYLRRKQTERSIDFTPKKVGQTS 1089

Qy        544 DFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKK-------LGNPKQPTNP 596
              . |          :|:    | |    : |        |      | | ::||          |      |
Db       1090 DITLVLGALLDTSHKFRNLDKDLCEKCAKCISMIG-VLDVTKHEFKRTTYSENEVYDLND 1148

Qy        597 FLEMIKFLL----ERIAPVHIDTESIS-----ALIKQVNKSIDGTADD-----EDEGVPT 642
               :: ||||:      :  : |       :|: |        ||: | :       | :  :           |  |
Db       1149 SVQTIKFLIWVINDILVPAFWQSENPSKQLFVALVIQESLKYCGLSSESWDMNHKELYPN 1208

Qy        643 DQAIRAGLELLKVLSFT--HPI--SFHSAETFESLLACLKMDDEKVAEA--------ALQ 690
              :   :      |       :| |    :|:    | :  |::::   :   ||             |            |
Db       1209 EAKL---WEKFNSVSKTTIYPLLSSLYLAQSWKEYVP-LKYPSNNFKEGYQIWVKRFTLD 1264

Qy        691 IFKNTG-----------SKIEEDFPHIRSALLPVLH----HKSKKGPP------------ 723
              : | ||                | | ||      : : |||  :          |::|| |
Db       1265 LLK-TGTTENHPGHVFSSLIREDDGSLSNFLLPYISLDIIIKAEKGTPYADILNGIIIEF 1323

Qy        724 -------------RQAKYAIHCIHAIFSSKE------TQFAQIFEPLH-----KSLDPSN 759
                           |        |  :||   |          |:| |  :  ||         |      :|
Db       1324 DSIFTCNLEGMNNLQVDSLRMCYESIFRVFEYCKKWATEFKQNYSKLHGTFIIKDTKTTN 1383

Qy        760 LEHLITPLVTIGHIALLAPDQFAAPWKSWVATFIVKDLLMNDRLP--------------- 804
              :                  ||  |:|        |             |||.   |
Db       1384 M--------------LLRIDEFLRTTPS--------DLLAQRSLETDSFERSALYLEQCY 1421

Qy        805 ----------GKKTTKLWVPDEEV---------------SPETMVKIQAIKMMVRW----- 835
                         |:       | :  ||:                         :   : ||: ::      |
Db       1422 RQNPHDKNQNGQLLKNLQITYEEIGDIDSLDGVLRTFATGNLVSKIEELQYSENWKLAQD 1481

Qy        836 ---LLGMKNNHSKSGTSTLR------LLTTILHSDGDLTEQGKIS-KPDMSRLRLAAGSA 885
                 :||  :: |:   |:      |:                   |  |: :         :    ||||   ||:
Db       1482 CFNVLGKFSDDPKTTTRMLKSMYDHQLYSQIISNSSFHSSDGKISLSPDVK--------- 1532

Qy        886 IVKLAQEPCYHEIITLEQYQLCALAINDE-CYQVRQVFAQKLH--KGLSRLRLPLEY--- 939
                                  | | :     | | |     |   : : :::     : :        : |:|
Db       1533 ----------------EWYSIGLEAANLEGNVQTLKNWVEQIESLRNIDDREVLLQYNIA 1576

Qy        940 MAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLL 999
               |: |:   :||:: :                | |:          :  :       |  |   :||  :     :      :| |
Db       1577 KALIAISNEDPLRTQ----------KYIHNSFRLIGTNFITSSKETTLLKKQNLLMKLHSL 1627

Qy       1000 AHDPDYV-KVQDIEQLKDVKECLWFVLEILMAKNENNSHAF-IRKMVENIKQTKDAQGPD 1057
               :|    ::     :|   :    |        |  :  :|  |     | :      :||    : :|    : |
Db       1628 -YDLSFLSSAKDKFEYKSNTTILDYRMERIGADFVPNHYILSMRKSFDQLKMNEQA---- 1682

Qy       1058 DAKMNEKLYTVCDVAMN----IIMSKSTTYSLE 1086
              || :  :   :|:  :| |       | |:|  : ||
Db       1683 DADLGKTFFTLAQLARNNARLDIASESLMHCLE 1715
```

RESULT 10
US-08-870-693-15
; Sequence 15, Application US/08870693
; Patent No. 5866338
; GENERAL INFORMATION:
; APPLICANT: Hartwell, Leland H.
; APPLICANT: Weinert, Ted A.
; APPLICANT: Plon, Sharon E.
; APPLICANT: Groudine, Mark T.
; TITLE OF INVENTION: Cell Cycle Checkpoint Genes
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Ave., Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,693
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/198,446
; FILING DATE: February 18, 1994
; APPLICATION NUMBER: PCT/US93/04458
; FILING DATE: May 12, 1993
; APPLICATION NUMBER: US 07/884,426
; FILING DATE: May 14, 1992
; APPLICATION NUMBER: US 07/882,051
; FILING DATE: May 12, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FHCR110798
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100
; TELEFAX: 206-224-0779
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2368 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: yeast MEC1 protein
US-08-870-693-15

```
  Query Match             2.5%;  Score 178.5;  DB 2;  Length 2368;
  Best Local Similarity   19.4%;  Pred. No. 5.5e-05;
  Matches  251;  Conservative  188;  Mismatches  477;  Indels  377;  Gaps   60;

Qy         68 FLKHPGKD--VRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQF 125
              || || |       |::  |   :  ||::    :  | |||  |       |  :  | |
Db        526 FLSHPNIDEFSESLLSGILFSLHRIFS----HFQPPKLTDGNGQINKSFK-LVQKCFMNS 580

Qy        126 NRYFYLLENIAWVKSYNI--CFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSS 183
              ||| ||          |  |    | : ||:     :    ||     : :         |   ::|   :
Db        581 NRYLRLL-------STRIIPLFNISDSHNSEDEHTATLIKFLQSQKLPVVKENLVIAWTQ 633

Qy        184 IICEGDTVSQELLDTVLVNLVPAHKN--------LNKQAYDLAKALLKRTAQAIEPYITT 235
              :        | | :: ||:|: |:      :         :   |  ::|| | |     |  : | :
Db        634 LTL---TTSNDVFDTLLLKLIDIFNSDDYSLRIMMTLQIKNMAKILKKTPYQLLSPILPV 690

Qy        236 FFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLLLS-----VLPQLEFKLKSNDNEERL 290
                 |  |||   :  |  ::||      :|     | :|      ::|     ||:
Db        691 LLRQ---LGKNLVE--RKVGFQNLIELLGYPSKTILDIFQRYIIPYAIIQYKSD------ 739

Qy        291 QVVKLLAKMFGAKDSELASQNKP-------------------------LWQCYLGR---F 322
              |:   :||:     |: | :| |                          |    :|  |  |
Db        740 -VLSEIAKIMCDGDTSLINQMKVNLLKKNSRQIFAVALVKHGLFSLDILETLFLNRAPTF 798

Qy        323 NDIHVPIRLECVK-FASHCLMNHPDLAKD-----------------LTEYLKVRSH---- 360
              :  ::     |    |   |       :      : ||                :|  :  |  :  |
Db        799 DKGYITAYLPDYKTLAEITKLYKNSVTKDASDSENANMILCSLRFLITNFEKDKRHGSKY 858
```

```
Qy        361 --------DPEEAIR---HDVIVSIVTAAKKDILLVNDHLLNFVRERTLDK-----RWRV 404
Db        859 KNINNWTDDQEQAFQKKLQDNILGIFQVFSSDIHDVEGRTTYYEKLRVINGISFLIIYAP 918

Qy        405 RKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRL--LVERIFA-- 460
Db        919 KKSIISALAQI--SICLQTGLG---LKEVRYEAFRCWHLLVRHLNDEELSTVIDSLIAFI 973

Qy        461 -----------QYMVPHNLETTERMKCLYYL----YATLDLNAVKALNEMWKCQNLLR-- 503
Db        974 LQKWSEFNGKLRNIVYSILDTLIKEKSDLILKLKPYTTLALVGKPELGILARDGQFARMV 1033

Qy        504 HQVKDLLDLIKQPKTDASVKAIFSKVMVITRNLPD-------------------PGKAQ 543
Db       1034 NKIRSTTDLI--PIFANNLKS--SNKYVINQNLDDIEVYLRRKQTERSIDFTPKKVGQTS 1089

Qy        544 DFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKK-------LGNPKQPTNP 596
Db       1090 DITLVLGALLDTSHKFRNLDKDLCEKCAKCISMIG-VLDVTKHEFKRTTYSENEVYDLND 1148

Qy        597 FLEMIKFLL----ERIAPVHIDTESIS-----ALIKQVNKSIDGTADD-----EDEGVPT 642
Db       1149 SVQTIKFLIWVINDILVPAFWQSENPSKQLFVALVIQESLKYCGLSSESWDMNHKELYPN 1208

Qy        643 DQAIRAGLELLKVLSFT--HPI--SFHSAETFESLLACLKMDDEKVAEA--------ALQ 690
Db       1209 EAKL---WEKFNSVSKTTIYPLLSSLYLAQSWKEYVP-LKYPSNNFKEGYQIWVKRFTLD 1264

Qy        691 IFKNTG----------SKIEEDFPHIRSALLPVLH----HKSKKGPP----------- 723
Db       1265 LLK-TGTTENHPGHVFSSLIREDDGSLSNFLLPYISLDIIIKAEKGTPYADILNGIIIEF 1323

Qy        724 -------------RQAKYAIHCIHAIFSSKE------TQFAQIFEPLH-----KSLDPSN 759
Db       1324 DSIFTCNLEGMNNLQVDSLRMCYESIFRVFEYCKKWATEFKQNYSKLHGTFIIKDTKTTN 1383

Qy        760 LEHLITPLVTIGHIALLAPDQFAAPWKSWVATFIVKDLLMNDRLP-------------- 804
Db       1384 M-------------LLRIDEFLRTTPS--------DLLAQRSLETDSFERSALYLEQCY 1421

Qy        805 ----------GKKTTKLWVPDEEV-------------SPETMVKIQAIKMMVRW----- 835
Db       1422 RQNPHDKNQNGQLLKNLQITYEEIGDIDSLDGVLRTFATGNLVSKIEELQYSENWKLAQD 1481

Qy        836 ---LLGMKNNHSKSGTSTLR------LLTTILHSDGDLTEQGKIS-KPDMSRLRLAAGSA 885
Db       1482 CFNVLGKFSDDPKTTTRMLKSMYDHQLYSQIISNSSFHSSDGKISLSPDVK--------- 1532

Qy        886 IVKLAQEPCYHEIITLEQYQLCALAINDE-CYQVRQVFAQKLH--KGLSRLRLPLEY--- 939
Db       1533 ---------------EWYSIGLEAANLEGNVQTLKNWVEQIESLRNIDDREVLLQYNIA 1576

Qy        940 MAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLL 999
Db       1577 KALIAISNEDPLRTQ---------KYIHNSFRLIGTNFITSSKETTLLKKQNLLMKLHSL 1627

Qy       1000 AHDPDYV-KVQDIEQLKDVKECLWFVLEILMAKNENNSHAF-IRKMVENIKQTKDAQGPD 1057
Db       1628 -YDLSFLSSAKDKFEYKSNTTILDYRMERIGADFVPNHYILSMRKSFDQLKMNEQA---- 1682

Qy       1058 DAKMNEKLYTVCDVAMN----IIMSKSTTYSLE 1086
Db       1683 DADLGKTFFTLAQLARNNARLDIASESLMHCLE 1715
```

RESULT 11
US-08-290-731C-2
; Sequence 2, Application US/08290731C
; Patent No. 5843646
; GENERAL INFORMATION:
; APPLICANT: BOWTELL, David Douglas Lawrence
; TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
; TITLE OF INVENTION: SON OF SEVENLESS (mSOS) GENE,
; TITLE OF INVENTION: AND mSOS POLYPEPTIDES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 PENNSYLVANIA AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,731C
; FILING DATE: 17-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU93/00068
; FILING DATE: 17-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PL0921/92
; FILING DATE: 17-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KIT, Gordon
; REGISTRATION NUMBER: 30,764
; REFERENCE/DOCKET NUMBER: Q-36066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1319 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-290-731C-2

```
  Query Match             2.4%;  Score 176;  DB 2;  Length 1319;
  Best Local Similarity   19.2%;  Pred. No. 3.4e-05;
  Matches  252;  Conservative  200;  Mismatches  494;  Indels  364;  Gaps   68;

Qy        189 DTVSQELLDTVLVNLVPAHKNLNKQA--YDLAKALLKRTAQAI----------EPYITT 235
Db        174 DMFHQDVEDINILSLTDEEPSTSGEQTYYDLVKAFMAEIRQYIRELNLIIKVFREPFVSN 233

Qy        236 FFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKL 295
Db        234 --------SKLFSSNDVENIFSRIVDIHELSVKLLGHIEDTVEMTDEGSPH-------PL 278

Qy        296 LAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCL--MNHPDLAKDLTE 353
Db        279 VGSCFEDLAEELA------FDPYESYARDILRP------GFHGHFLSQLSKPGAAL---- 322

Qy        354 YLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDH---LLNFVRERTLDKRWR-VRKEAM 409
Db        323 YLQSIGEGFKEAVQY-VLPRLLLAPVYHCL----HYFELLKQLEEKSEDQEDKECMKQAI 377

Qy        410 MGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVE---RIFAQYMVPH 466
Db        378 TALLNV------QSGMEKICSKSLA----------------KRRLSESACRFYSQQMKGK 415

Qy        467 NLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLD-----LIKQPKTDAS 521
Db        416 QL-------------------AIKKMNEIQK--NIDGWEGKDIGQCCNEFIMEGTLTRVG 454

Qy        522 VK-----AIFSKVMVITRN------LPDPGKAQ------DFMKKFTQVLEDDE--KIRKQ 562
Db        455 AKHERHIFLFDGLMICCKSNHGQPRLPGASSAEYRLKEKFFMRK-VQINDKDDTSEYKHA 513
```

```
Qy        563 LEVLV----SPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMI----KFLLERIAPVHID 614
               |:::    |   | | ||                :  |    :|    :  |||:  | :
Db        514 FEIILKDGNSVIFSAKSAE-------------EKNNWMAALISLQYRSTLERMLDVTVL 559

Qy        615 TESISALIKQVNKSIDGTAD-DEDEGVPTDQAI--RAGLELLK---VLSFTHPISFHS-- 666
              |     ::  :  :    |: | :| :  :: :  :||: ::|   ||      :::|
Db        560 QEEKEEQMRLPSAEVYRFAEPDSEENILFEENVQPKAGIPIIKAGTVLKLIERLTYHMYA 619

Qy        667 ----AETF---------ESLLACL----KMDDEKVAEAALQIFKNTGSKIEEDFPHIRS 708
                  ||           : ||: |    :: : :  ||      :|      :    :    |
Db        620 DPNFVRTFLTTYRSFCRPQELLSLLIERFEIPEPEPTEADRIAIENGDQPLSAELKRFRK 679

Qy        709 ALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLV 768
                :          | | :    | | :     |  |   ||       |    |: :    :
Db        680 EYI---------QPVQLRVLNVCRHWV----EHHFYD-FE-----RDADLLQRMEEFIG 719

Qy        769 TIGHIALLAPDQFAAPWKSWVAT---FIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVK 825
              |:   |:            | ||  :       | :  :   |   ||    |          : || |
Db        720 TVRGKAM---------KKWVESITKIIQRKKIARDNGPGHNIT------FQSSPPT--- 760

Qy        826 IQAIKMMVRWLLGMKNNHSKSG-TSTLRLLT---------TILHSD----------GD 863
                  | | :         |:  |     |||           |:| ||               |
Db        761 -------VEWHI------SRPGHIETFDLLTLHPIEIARQLTLLESDLYRAVQPSELVGS 807

Qy        864 L--TEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQV 921
               :    |   :|: |::   |::     :    :   |   |   |      ||:            :|:         :: ||
Db        808 VWTKEDKEINSPNL--LKMIRHTTNLTLWFEKCIVETENLEE----RVAVVSRIIEILQV 861

Qy        922 FAQKLHKGLSRLRLPLEYMAICALCAKDPV-------KERRAHARQCLVKNINVRREYLK 974
              | :        |:      ||    :  :         ||           ::  :  ::  |  :     :  :: |
Db        862 FQE-----LNNFNGVLE---VVSAMNSSPVYRLDHTFEQIPSRQKKILEEAHELSEDHYK 913

Qy        975 QHAAVSEKLLSLLPEYVVPYTIHLL------AHDPDYVKVQDIEQLKDVKECLWF----- 1023
              :: |    || |: |   |    : |:|          :|: ::        :  || : |
Db        914 KYLA---KLRSINPPCVPFFGIYLTNILKTEEGNPEVLR-------RHGKELINFSKRRR 963

Qy       1024 VLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTY 1083
              | ||    :   :   ::  :||:   :         |  ||  :|        : : :||
Db        964 VAEITGEIQQYQNQPYCLRVEPDIKRFFENLNPMGNSM-EKEFT------DYLFNKSLEI 1016

Qy       1084 SLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPL---SS 1140
                 ||    || ||                | |   |           |:          |  :
Db       1017 EPRHPKP--LP-RF---------PKKYSYPLKSPGVRPSNPRP----GTMRHPTPLQQE 1059

Qy       1141 AGKQSQTKSSRMETVSNASSSSNPS---SPGRIKGRLDSSEM------DHSENEDYTMSS 1191
                 | | ::    || | ||: ::|     :|      |     ::::          ||| :  :: |:
Db       1060 PRKISYSRIPESETESTASAPNSPRTPLTPPPASGTSSNTDVCSVFDSDHSASPFHSRSA 1119

Qy       1192 PLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKL---GMDDLTKLVQEQ------K 1242
                  :      |   |   |    :   |  :  |          |:     :|     :      |
Db       1120 SVSSISLSKGTDEVPVPPPVPPRRRPESAPAESSPSKIMSKHLDSPPAIPPRQPTSKAYS 1179

Qy       1243 PKGSQRSRKRGHTASESDEQQWPEEKRLKEDILENEDEQNSPPKKGKRG---------- 1291
              |:  |    |          ||       | |        |:   :          ||   ||:
Db       1180 PRYSISDRTSISDPPESPPLLPPREPVRTPDVFSSSPLHLQPPPLGKKSDHGNAFFPNSP 1239

Qy       1292 ---RPPKPLGGGTPKEEPTMK-------TSKKGSKKKSGPPAPEEEEEEE 1331
                    || |      ||       |              |  :        :||| :  |
Db       1240 SPFTPPPP---QTPSPHGTRRHLPSPPLTQEMDLHSIAGPPVPPRQSTSQ 1286
```

RESULT 12
US-08-290-731C-6
; Sequence 6, Application US/08290731C
; Patent No. 5843646
; GENERAL INFORMATION:
; APPLICANT: BOWTELL, David Douglas Lawrence
; TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
; TITLE OF INVENTION: SON OF SEVENLESS (mSOS) GENE,
; TITLE OF INVENTION: AND mSOS POLYPEPTIDES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 PENNSYLVANIA AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,731C
; FILING DATE: 17-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU93/00068
; FILING DATE: 17-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PL0921/92
; FILING DATE: 17-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KIT, Gordon
; REGISTRATION NUMBER: 30,764
; REFERENCE/DOCKET NUMBER: Q-36066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1336 amino acids
; TYPE: amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-290-731C-6

```
  Query Match             2.4%;  Score 176;  DB 2;  Length 1336;
  Best Local Similarity   19.2%;  Pred. No. 3.4e-05;
  Matches  252;  Conservative  200;  Mismatches  494;  Indels  364;  Gaps   68;

Qy        189 DTVSQELLDTVLVNLVPAHKNLNKQA--YDLAKALLKRTAQAI----------EPYITT 235
              |   |::  |   :::|       : :  :   ||| ||  :      | |               ||:::
Db        191 DMFHQDVEDINILSLTDEEPSTSGEQTYYDLVKAFMAEIRQYIRELNLIIKVFREPFVSN 250

Qy        236 FFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKL 295
                          |   |:   |::|  |:::: :   ||   :   :|   : :  :         |
Db        251 --------SKLFSSNDVENIFSRIVDIHELSVKLLGHIEDTVEMTDEGSPH-------PL 295

Qy        296 LAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCL--MNHPDLAKDLTE 353
               :    |     |||        :    |        ||  |          |  |  |  :: |  |
Db        296 VGSCFEDLAEELA------FDPYESYARDILRP------GFHGHFLSQLSKPGAAL---- 339

Qy        354 YLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDH---LLNFVRERTLDKRWR-VRKEAM 409
              ||:       :||:::  |:    :: |        |       |     ||   :  |::  |:   :     |:|:
Db        340 YLQSIGEGFKEAVQY-VLPRLLLAPVYHCL----HYFELLKQLEEKSEDQEDKECMKQAI 394

Qy        410 MGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVE---RIFAQYMVPH 466
              |      :         ||    | :|                      |  | |      | ::| |
Db        395 TALLNV------QSGMEKICSKSLA---------------KRRLSESACRFYSQQMKGK 432

Qy        467 NLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLD-----LIKQPKTDAS 521
               |                     |:|  :||: |   |:    :  ||:              ::: |
Db        433 QL------------------AIKKMNEIQK--NIDGWEGKDIGQCCNEFIMEGTLTRVG 471

Qy        522 VK-----AIFSKVMVITRN------LPDPGKAQ------DFMKKFTQVLEDDE--KIRKQ 562
               |         :|  :|:  ::       ||       |:                 ||:|  |: : |:  : :
Db        472 AKHERHIFLFDGLMICCKSNHGQPRLPGASSAEYRLKEKFFMRK-VQINDKDDTSEYKHA 530

Qy        563 LEVLV----SPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMI----KFLLERIAPVHID 614
               |:::    |   | | ||                :  |    :|    :  |||:  | :
Db        531 FEIILKDGNSVIFSAKSAE-------------EKNNWMAALISLQYRSTLERMLDVTVL 576

Qy        615 TESISALIKQVNKSIDGTAD-DEDEGVPTDQAI--RAGLELLK---VLSFTHPISFHS-- 666
```

```
Db        577 QEEKEEQMRLPSAEVYRFAEPDSEENILFEENVQPKAGIPIIKAGTVLKLIERLTYHMYA 636

Qy        667 ----AETF----------ESLLACL----KMDDEKVAEAALQIFKNTGSKIEEDFPHIRS 708

Db        637 DPNFVRTFLTTYRSFCRPQELLSLLIERFEIPEPEPTEADRIAIENGDQPLSAELKRFRK 696

Qy        709 ALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLV 768

Db        697 EYI---------QPVQLRVLNVCRHWV----EHHFYD-FE-----RDADLLQRMEEFIG 736

Qy        769 TIGHIALLAPDQFAAPWKSWVAT---FIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVK 825

Db        737 TVRGKAM---------KKWVESITKIIQRKKIARDNGPGHNIT------FQSSPPT--- 777

Qy        826 IQAIKMMVRWLLGMKNNHSKSG-TSTLRLLT----------TILHSD-----------GD 863

Db        778 -------VEWHI------SRPGHIETFDLLTLHPIEIARQLTLLESDLYRAVQPSELVGS 824

Qy        864 L--TEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQV 921

Db        825 VWTKEDKEINSPNL--LKMIRHTTNLTLWFEKCIVETENLEE----RVAVVSRIIEILQV 878

Qy        922 FAQKLHKGLSRLRLPLEYMAICALCAKDPV-------KERRAHARQCLVKNINVRREYLK 974

Db        879 FQE-----LNNFNGVLE---VVSAMNSSPVYRLDHTFEQIPSRQKKILEEAHELSEDHYK 930

Qy        975 QHAAVSEKLLSLLPEYVVPYTIHLL------AHDPDYVKVQDIEQLKDVKECLWF----- 1023

Db        931 KYLA---KLRSINPPCVPFFGIYLTNILKTEEGNPEVLR-------RHGKELINFSKRRR 980

Qy       1024 VLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTY 1083

Db        981 VAEITGEIQQYQNQPYCLRVEPDIKRFFENLNPMGNSM-EKEFT------DYLFNKSLEI 1033

Qy       1084 SLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPL---SS 1140

Db       1034 EPRHPKP--LP-RF----------PKKYSYPLKSPGVRPSNPRP----GTMRHPTPLQQE 1076

Qy       1141 AGKQSQTKSSRMETVSNASSSSNPS---SPGRIKGRLDSSEM------DHSENEDYTMSS 1191

Db       1077 PRKISYSRIPESETESTASAPNSPRTPLTPPPASGTSSNTDVCSVFDSDHSASPFHSRSA 1136

Qy       1192 PLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKL---GMDDLTKLVQEQ------K 1242

Db       1137 SVSSISLSKGTDEVPVPPPVPPRRRPESAPAESSPSKIMSKHLDSPPAIPPRQPTSKAYS 1196

Qy       1243 PKGSQRSRKRGHTASESDEQQWPEEKRLKEDILENEDEQNSPPKKGKRG----------- 1291

Db       1197 PRYSISDRTSISDPPESPPLLPPREPVRTPDVFSSSPLHLQPPPLGKKSDHGNAFFPNSP 1256

Qy       1292 ---RPPKPLGGGTPKEEPTMK-------TSKKGSKKKSGPPAPEEEEEEE 1331

Db       1257 SPFTPPPP---QTPSPHGTRRHLPSPPLTQEMDLHSIAGPPVPPRQSTSQ 1303
```

RESULT 13
US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Polypeptide
US-08-973-462-8

Query Match 2.4%; Score 174; DB 3; Length 1786;
Best Local Similarity 17.0%; Pred. No. 8.1e-05;
Matches 203; Conservative 192; Mismatches 398; Indels 402; Gaps 48;

```
Qy        101 DKLKDIFM---------FITRQLKGLEDTKSPQFNRYFYLLENIA--WVKSYNICFELED 149

Db        838 EEIKDSILNEIEEVKENVVTTILENVEETTAESVTTFSNILEEIQENTITNDTIEEKLEE 897

Qy        150 SNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSSIICEGDTVSQELLDTV-------LVN 202

Db        898 LHE------NVLSAALENTQSEEEKKEVIDVIEEV---KEEVATTLIETVEQAEEKSANT 948

Qy        203 LVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTFFNQVLMLGKTSISDLSEHVFDLILEL 262

Db        949 ITEIFENLEENAVESN----ENVAENLEKLNETVFNTVL--------------DKVEET 989

Qy        263 YNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRF 322

Db        990 VEISGESLENNEMDKAFFSEIFDNVKGIQ-ENLLTGMFRSIETSIVIQSEEKVD-----L 1043

Qy        323 NDIHVPIRLECVKFASHCLMNHPD-------LAKDLTEYLKVRSHDPEEAIRHDVIVSIV 375

Db       1044 NENVVSSILDNIENMKEGLLNKLENISSTEGVQETVTEH-----------VEQNVYVDV- 1091

Qy        376 TAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAW 435

Db       1092 -----DVPAMKDQFLGILNEA------GGLKEMFFNLEDVFK------------------ 1122

Qy        436 IKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEM 495

Db       1123 -----------SESDVITVEEIKDE---PVQKE-------------VEKETVSIIEEM 1153

Qy        496 WKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLED 555

Db       1154 -------EENIVDVLEEEKEDLTDKMIDAVEESIEISSDSKEETESIKDKEKDVSLVVEE 1206

Qy        556 ------DEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPF-LEMIKFLLERI 608

Db       1207 VQDNDMDESVEKVLELKNMEEELMKDAVE-INDITSKLIEETQELNEVEADLIK------ 1259

Qy        609 APVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAE 668

Db       1260 -----DMEKLKELEKALSEDSKEIIDAKDD------------TLEKVIEEEHDIT----T 1298

Qy        669 TFESLLACLKMDDEKVAEAA-----------LQIFKNTGSKIEEDFPHIRSALLPVLHH 716

Db       1299 TLDEVVELKDVEEDKIEKVSDLKDLEEDILKEVKEIKELESEILEDYKELKTIETDILEE 1358

Qy        717 KSKKGPPRQAKYAIHC---------IHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPL 767

Db       1359 KKEIEKDHFEKFEEEAEEIKDLEADILKEVSSLEVEEEKKLEEVHELKE--EVEHIIS-- 1414

Qy        768 VTIGHIALLAPDQF--AAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVK 825

Db       1415 -GDAHIKGLEEDDLEEVDDLKGSILDMLKGDMELGDM--DKESL------EDVTTKLGER 1465

Qy        826 IQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSA 885

Db       1466 VESLKDVLSSALGMDEEQMKTR-------------------------------------- 1487

Qy        886 IVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICAL 945

Db       1488 --KKAQRPKLEEVLLKE-------------EVKEEPKKKITK--KKVRFDIK------- 1522

Qy        946 CAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDY 1005
```

```
                 |:|  |           :  : :: | :::   | |                    :  |  :
Db        1523 -DKEPKDE---------IVEVEMKDEDIEED--VEED---------------IEEDIEE 1554

Qy        1006 VKVQDIEQ--LKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNE 1063
               ||:||::    :|: |     |:::::  |              | :| :| |       |: :
Db        1555 DKVEDIDEDIDEDIGEDKDEVIDLIVQK---------EKRIEKVKAKK-------KKLEK 1598

Qy        1064 K-------LYTVCDVAMNII------MSKSTTYSLESPKD------------------- 1090
               |        |    |   |   :         : |   : :|||  |
Db        1599 KVEEGVSGLKKHVDEVMKYVQKIDKEVDKEVSKALESKNDVTNVLKQNQDFFSKVKNFVK 1658

Qy        1091 -------PVLPA---------RFFT-----------------QPDKN------------ 1104
                      | :  |           |||                  : ||
Db        1659 KYKVFAAPFISAVAAFASYVVGFFTFSLFSSCVTIASSTYLLSKVDKTINKNKERPFYSF 1718

Qy        1105 ----FSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETV 1155
                    | | |:|| :|| |:  | |  ||:   ||      |     |: :   |
Db        1719 VFDIFKNLKHYL-QQMKEKFS--KEKNNNVIEVTNKAEKKGNVQVTNKTEKTTKV 1770
```

RESULT 14
US-08-588-985-2
; Sequence 2, Application US/08588985
; Patent No. 5777094
; GENERAL INFORMATION:
;   APPLICANT: Michiyuki MATSUDA et al.
;   TITLE OF INVENTION: cDNA OF DOCK180 GENE AND DOCK180 PROTEIN
;   NUMBER OF SEQUENCES: 2
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Wenderoth, Lind & Ponack
;   STREET: 805 Fifteenth Street, N.W., #700
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: U.S.A.
;   ZIP: 20005
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: MS-DOS
;   SOFTWARE: Wordperfect 5.1
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/588,985
;   FILING DATE: January 19, 1996
;   CLASSIFICATION: 536
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Warren M. Cheek, Jr.
;   REGISTRATION NUMBER: 33,367
;   REFERENCE/DOCKET NUMBER:
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-371-8850
;   TELEFAX:
;   TELEX:
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1865 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   HYPOTHETICAL:
;   ANTI-SENSE:
;   FRAGMENT TYPE:
;   ORIGINAL SOURCE: spleen cell of homo sapiens
US-08-588-985-2

```
  Query Match             2.4%;  Score 172.5;  DB 1;  Length 1865;
  Best Local Similarity   17.6%;  Pred. No. 0.00012;
  Matches  258;  Conservative  212;  Mismatches  441;  Indels  553;  Gaps   72;

Qy           6 TRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLALHLAS 65
               |  || |:  :  |  | |:  : | |                 :  |      |   |:|
Db         803 TIVNDVKLVFDP--KELS-----------KMFTEFILNVPMGLLTIQKLYCLIEI-VHS 847

Qy          66 DFFLKHPGKDVRL---------------LVACC--LADIFRI-YAPEAPYTSPDKLKDI 106
               | | :|   ::: |              | |||  |: |  : |  :    |     : :
Db         848 DLFTQHDCREILLPMMTDQLKYHLERQEDLEACCQLLSHILEVLYRKDVGPTQ----RHV 903

Qy         107 FMFITRQLKGLEDTKSPQFNRYFYLLEN-IAWVKSYNICFELEDSNEIFTQLYRTLFSVI 165
                : : : |: :  |         |    |: | :|: : :  |  ::|| :  : | :|
Db         904 QIIMEKLLRTVNRTVI-SMGRDSELIGNFVACMTA--ILRQMEDYH--YAHLIKTF---- 954

Qy         166 NNGHNQKVHMHMVD-LMSSIICEGDTVSQEL--LDTVLVNLVPAHKNLNKQAYDLAKALL 222
                    |:    :|| || : |    : : : :   | |::|:|            |  |
Db         955 -----GKMRTDVVDFLMETFIMFKNLIGKNVYPFDWVIMNMVQ-----------NKVFL 997

Qy         223 KRTAQAIEPYITTFFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLK 282
               :     || |        ||| |   :    :   |:  |:|:|    || :: |  |   :|:
Db         998 R----AINQYAD-------MLNKKFLDQAN---FE--LQLWNNYFHLAVAFLTQESLQLE 1041

Qy         283 SNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLM 342
                : : :|     |:| | :|    ::  : : :|   ||:              :||
Db        1042 NFSSAKR---AKILNK-YGDMRRQIGFEIRDMWY-NLGQHK---------IKFI----- 1081

Qy         343 NHPDLAKDLTEYLKVRSHDPEEAIRHDVI------------VSIVTAAKKDILLVNDHL 389
                 |::    :  |     :   || :|  |                        : :|:  ||
Db        1082 --PEMVGPILEMTLI----PETELRKATIPIFFDMMQCEFHSTRSFQMFENEIITKLDHE 1135

Qy         390 LNFVRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSI 449
                :     |     |::::|   :  :   |   :  :|:    :   |:    |  :
Db        1136 VEGGRG---DEQYKVLFDKI--LLEHCRKHKYLAKTGETFVKLVV-------------- 1175

Qy         450 DDRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDL 509
                    |:||:      : |:        | |        :| :      |: :  :  :|:  |  |
Db        1176 ---RLMERLLDYRTIMHDENKENRMSC--------TVNVLNFYKEIEREEMYIRYLYK-L 1223

Qy         510 LDLIKQPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSP 569
                 || |:     |     :|  :: ::                         ::|:   |  :
Db        1224 CDLHKE--CDNYTEAAYTLLL-----------------HAKLLKWSEDV--------- 1253

Qy         570 TCSCKQAEGCVREITKKLGNPKQPTNP-------FLEMIKFLLERIAPVHIDTESISALI 622
                          ||  :|:: |    |  |           : |:| :
Db        1254 ---------CVAHLTQRDG--YQATTQGQLKEQLYQEIIHYF----------------- 1284

Qy         623 KQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDE 682
                                    |:|     ::|| | ||                 || :|: :     | |
Db        1285 -------------DKGKMWEEAIALGKEL-------------AEQYENEM----FDYE 1312

Qy         683 KVAE---AALQIFKNTGSKIE-----------EDFPHIRSALLPVLHHKSKKGPPRQAK 727
               :::|       | ::|      |                  : ||          |   :: |:   ||:
Db        1313 QLSELLKKQAQFYENIVKVIRPKPDYFAVGYYGQGFPTFLRG--KVFIYRGKEYEPRED- 1369

Qy         728 YAIHCIHAIFSSK-ETQFAQIFEPLHKSLDPSNLEH---------LITPLVTIGHIALLA 777
                            | ::  |||      :       |        ::::                 :  |  :  :
Db        1370 ---------FEARLLTQFPNAEKMKTTSPPGDDIKNSPGQYIQCFTVKPKLDL------- 1413

Qy         778 PDQFAAPWKSWVATF----------IVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQ 827
               | :|  |        : :|                 :  :    :: |   :    :|:   |        |  |:
Db        1414 PPKFHRPVSEQIVSFYRVNEVQRFEYSRPIRKGEKNPDNEFANMWI--ERTIYTTAYKLP 1471

Qy         828 AIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIV 887
                 |   :||                               |     :   ::|  |   |
Db        1472 GI---LRWF----------------------------EVKSVFMVEISPLENA------ 1493

Qy         888 KLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCA 947
                            :|   ||        ||                   :  :   |    ||:  :::
Db        1494 ------------IETMQLTNDKIN----------SMVQQHLDDPSLPINPLSMLLNGI 1529

Qy         948 KDP-VKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYV 1006
                 || |        |:  :     :        ||::|      ||
Db        1530 VDPAVMGGFANYEKAFFTD-----RYLQEHPEAHEK------------------------ 1560

Qy        1007 KVQDIEQLKDVKECLW---FVLEILMAKNENNSHA---FIRKMVENIKQTKDAQGPDDAK 1060
                   ||:|||:    |      |:  |  :        :   |    |    :|        || |:
```

```
Db       1561 ----IEKLKDL--IAWQIPFLAEGIRIHGDKVTEALRPFHERMEACFKQLKEKV------ 1608

Qy       1061 MNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFF 1120
                ||  |  |       |: |          |:  |    |  || |
Db       1609 --EKEYGV-----RIMPSSLDDRRGSRPRSMV---RSFTMPSS---------------- 1641

Qy       1121 TPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMD 1180
                             ::|||  |           :||:  ||  |  || ||
Db       1642 -------------SRPLSVA-----------SVSSLSSDSTPSRPG------------ 1663

Qy       1181 HSENEDYTMSSPLPGK----KSDKRDDSDLVRSELEKPRGR------------KKTPVT 1223
                 :: :  :     ||  |           || |   || :  : :| :            | |  ::
Db       1664 ---SDGFALEPLLPKKMHSRSQDKLDKDDLEKEKKDKKKEKRNSKHQEIFEKEFKPTDIS 1720

Qy       1224 -EQEEKLGMDDLTKLVQEQKPK-------GSQRSRKRGHTASESDEQQWPEEKRLK---- 1271
               :|  |  :  :  :      ::  |:||       ||:|          ::  |  :    |    |  |
Db       1721 LQQSEAVILSETISPLRPQRPKSQVMNVIGSERRFSVSPSSPSSQQTPPPVTPRAKLSFS 1780

Qy       1272 ------------------------------EDILENEDEQNS----PPKKGKRGRPPKP 1296
                                            :::||:|    |      ||    :|   || |
Db       1781 MQSSLELNGMTGADVADVPPPLPLKGSVADYGNLMENQDLLGSPTPPPPPPHQRHLPP-P 1839

Qy       1297 LGGGTPKEEPTMKTSKKGSKKKSG 1320
                |    ||    |    ||::|  :     ||
Db       1840 LPSKTPPPPPP-KTTRKQTSVDSG 1862
```

RESULT 15

US-08-971-988-2

```
; Sequence 2, Application US/08971988
; Patent No. 5786461
;  GENERAL INFORMATION:
;    APPLICANT:  Michiyuki MATSUDA et al.
;    TITLE OF INVENTION:  cDNA OF DOCK180 GENE AND DOCK180 PROTEIN
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 500 kb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/971,988
;      FILING DATE:  17-NOV-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/588,985
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-8850
;      TELEFAX:
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1865 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:
;    ANTI-SENSE:
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:  spleen cell of homo sapiens
```

US-08-971-988-2

```
  Query Match             2.4%;  Score 172.5;  DB 1;  Length 1865;
  Best Local Similarity   17.6%;  Pred. No. 0.00012;
  Matches  258;  Conservative  212;  Mismatches  441;  Indels  553;  Gaps   72;

Qy          6 TRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLALHLAS 65
Db        803 TIVNDVKLVFDP--KELS-----------KMFTEFILNVPMGLLTIQKLYCLIEI-VHS 847

Qy         66 DFFLKHPGKDVRL---------------LVACC--LADIFRI-YAPEAPYTSPDKLKDI 106
Db        848 DLFTQHDCREILLPMMTDQLKYHLERQEDLEACCQLLSHILEVLYRKDVGPTQ----RHV 903

Qy        107 FMFITRQLKGLEDTKSPQFNRYFYLLEN-IAWVKSYNICFELEDSNEIFTQLYRTLFSVI 165
Db        904 QIIMEKLLRTVNRTVI-SMGRDSELIGNFVACMTA--ILRQMEDYH--YAHLIKTF---- 954

Qy        166 NNGHNQKVHMHMVD-LMSSIICEGDTVSQEL--LDTVLVNLVPAHKNLNKQAYDLAKALL 222
Db        955 -----GKMRTDVVDFLMETFIMFKNLIGKNVYPFDWVIMNMVQ-----------NKVFL 997

Qy        223 KRTAQAIEPYITTFFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLK 282
Db        998 R----AINQYAD------MLNKKFLDQAN---FE--LQLWNNYFHLAVAFLTQESLQLE 1041

Qy        283 SNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLM 342
Db       1042 NFSSAKR---AKILNK-YGDMRRQIGFEIRDMWY-NLGQHK----------IKFI----- 1081

Qy        343 NHPDLAKDLTEYLKVRSHDPEEAIRHDVI------------VSIVTAAKKDILLVNDHL 389
Db       1082 --PEMVGPILEMTLI----PETELRKATIPIFFDMMQCEFHSTRSFQMFENEIITKLDHE 1135

Qy        390 LNFVRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSI 449
Db       1136 VEGGRG---DEQYKVLFDKI--LLEHCRKHKYLAKTGETFVKLVV-------------- 1175

Qy        450 DDRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDL 509
Db       1176 ---RLMERLLDYRTIMHDENKENRMSC-------TVNVLNFYKEIEREEMYIRYLYK-L 1223

Qy        510 LDLIKQPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSP 569
Db       1224 CDLHKE--CDNYTEAAYTLLL-----------------HAKLLKWSEDV--------- 1253

Qy        570 TCSCKQAEGCVREITKKLGNPKQPTNP-------FLEMIKFLLERIAPVHIDTESISALI 622
Db       1254 ---------CVAHLTQRDG--YQATTQGQLKEQLYQEIIHYF----------------- 1284

Qy        623 KQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDE 682
Db       1285 -------------DKGKMWEEAIALGKEL-------------AEQYENEM----FDYE 1312

Qy        683 KVAE---AALQIFKNTGSKIE-----------EDFPHIRSALLPVLHHKSKKGPPRQAK 727
Db       1313 QLSELLKKQAQFYENIVKVIRPKPDYFAVGYYGQGFPTFLRG--KVFIYRGKEYEPRED- 1369

Qy        728 YAIHCIHAIFSSK-ETQFAQIFEPLHKSLDPSNLEH---------LITPLVTIGHIALLA 777
Db       1370 ---------FEARLLTQFPNAEKMKTTSPPGDDIKNSPGQYIQCFTVKPKLDL------- 1413

Qy        778 PDQFAAPWKSWVATF---------IVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQ 827
Db       1414 PPKFHRPVSEQIVSFYRVNEVQRFEYSRPIRKGEKNPDNEFANMWI--ERTIYTTAYKLP 1471

Qy        828 AIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIV 887
Db       1472 GI---LRWF----------------------------EVKSVFMVEISPLENA------ 1493

Qy        888 KLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCA 947
Db       1494 -------------IETMQLTNDKIN-----------SMVQQHLDDPSLPINPLSMLLNGI 1529
```

```
Qy        948 KDP-VKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYV 1006
               || |    |:  :    :      ||::|   ||
Db       1530 VDPAVMGGFANYEKAFFTD-----RYLQEHPEAHEK------------------------ 1560

Qy       1007 KVQDIEQLKDVKECLW---FVLEILMAKNENNSHA---FIRKMVENIKQTKDAQGPDDAK 1060
                  ||:|||:    |    |: |  :    :  : |    |  :|    || |:
Db       1561 ----IEKLKDL--IAWQIPFLAEGIRIHGDKVTEALRPFHERMEACFKQLKEKV------ 1608

Qy       1061 MNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFF 1120
                || | |        |: |          |:  |    | || |
Db       1609 --EKEYGV-----RIMPSSLDDRRGSRPRSMV---RSFTMPSS----------------- 1641

Qy       1121 TPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMD 1180
                             ::||| |           :||: || | || ||
Db       1642 --------------SRPLSVA------------SVSSLSSDSTPSRPG------------ 1663

Qy       1181 HSENEDYTMSSPLPGK----KSDKRDDSDLVRSELEKPRGR-------------KKTPVT 1223
                 :: :  :   || |       || |  || : : :| : :             | |  ::
Db       1664 ---SDGFALEPLLPKKMHSRSQDKLDKDDLEKEKKDKKKEKRNSKHQEIFEKEFKPTDIS 1720

Qy       1224 -EQEEKLGMDDLTKLVQEQKPK-------GSQRSRKRGHTASESDEQQWPEEKRLK---- 1271
               :| | : : :     :: |:||       ||:|        ::  | :    |    | |
Db       1721 LQQSEAVILSETISPLRPQRPKSQVMNVIGSERRFSVPSSPSSQQTPPPVTPRAKLSFS 1780

Qy       1272 ------------------------------EDILENEDEQNS----PPKKGKRGRPPKP 1296
                                            :::||:|    |     ||    :|  || |
Db       1781 MQSSLELNGMTGADVADVPPPLPLKGSVADYGNLMENQDLLGSPTPPPPPPHQRHLPP-P 1839

Qy       1297 LGGGTPKEEPTMKTSKKGSKKKSG 1320
              |   ||   |  ||::| :   ||
Db       1840 LPSKTPPPPPP-KTTRKQTSVDSG 1862
```

Search completed: September 24, 2003, 20:21:59
Job time : 35 secs

```
OM protein - protein search, using sw model

Run on:         September 24, 2003, 16:28:13 ; Search time 59 Seconds
                                              (without alignments)
                                              3742.181 Million cell updates/sec

Title:          US-09-512-581B-2
Perfect score:  7193
Sequence:       1 MAHSKTRTNDGKITYPPGVK..........QKGRGRPSKTPSPSQPKKNV 1391

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters:      1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      A_Geneseq_19Jun03:*
                1:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
                2:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
                3:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
                4:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
                5:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
                6:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
                7:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
                8:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
                9:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
                10:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
                11:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
                12:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
                13:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
                14:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
                15:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
                16:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
                17:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
                18:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
                19:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
                20:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
                21:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
                22:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
                23:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
                24:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7193	100.0	1391	21	AAY94702	Human androgen shu
2	4642.5	64.5	1297	24	ABG72803	Human tumour supre
3	2999.5	41.7	919	23	ABP64685	Human protein SEQ
4	2715.5	37.8	854	21	AAB01382	Neuron-associated
5	1982.5	27.6	1218	22	ABB67869	Drosophila melanog
6	839	11.7	217	22	AAU27827	Human full-length
7	758.5	10.5	363	21	AAB58462	Lung cancer associ
8	709	9.9	182	23	ABB84560	Human euchromosome
9	640.5	8.9	333	22	AAB94946	Human protein sequ
10	599	8.3	165	22	AAU01759	Human secreted pro
11	591	8.2	147	23	ABB89260	Human polypeptide
12	513	7.1	101	21	AAY65247	Human 5' EST relat
13	467	6.5	92	21	AAG02811	Human secreted pro
14	276	3.8	111	22	AAU27999	Human contig polyp
15	247.5	3.4	233	22	AAB95449	Human protein sequ
16	232	3.2	2688	22	AAM40883	Human polypeptide
17	231.5	3.2	2835	23	AAU75100	Novel signal pathw
18	231	3.2	2663	22	AAM39097	Human polypeptide
19	229	3.2	2633	22	ABG06505	Novel human diagno
20	228.5	3.2	6815	22	ABB66811	Drosophila melanog
21	225	3.1	1743	19	AAW98879	H. pylori GHPO 175
22	218.5	3.0	1026	22	AAM78825	Human protein SEQ
23	218	3.0	1033	22	AAM79809	Human protein SEQ
24	210.5	2.9	2047	23	AAU75883	Human adhesion mol
25	208.5	2.9	5303	22	ABB67866	Drosophila melanog
26	208	2.9	364	22	ABU52998	Human testes-deriv
27	208	2.9	628	22	ABU52994	Human testes-deriv
28	208	2.9	1183	22	ABB58769	Drosophila melanog
29	207	2.9	1428	22	ABB68557	Drosophila melanog
30	206	2.9	2243	22	AAB84884	Murine protein, SE
31	206	2.9	2400	22	ABG20278	Novel human diagno
32	205.5	2.9	2230	24	ABU07445	Protein differenti
33	205	2.8	2349	24	AAO16359	Human translocated
34	205	2.8	2415	22	ABG20279	Novel human diagno
35	203.5	2.8	1881	23	ABP73809	Candida albicans e
36	203	2.8	622	22	ABU52993	Human testes-deriv
37	201.5	2.8	3248	17	AAR99795	Kinetochore protei
38	200	2.8	1278	22	ABB71882	Drosophila melanog
39	199.5	2.8	699	22	AAM79156	Human protein SEQ
40	199.5	2.8	718	22	AAM80140	Human protein SEQ
41	197.5	2.7	1530	22	AAB84882	Human protein, SEQ
42	197.5	2.7	2266	22	AAB84885	Human protein, SEQ
43	195.5	2.7	1150	22	ABB59129	Drosophila melanog
44	195	2.7	2519	22	ABG16636	Novel human diagno
45	194.5	2.7	2482	16	AAR72826	Human mitosin. Ho

ALIGNMENTS

```
RESULT 1
AAY94702
ID   AAY94702 standard; Protein; 1391 AA.
XX
AC   AAY94702;
XX
DT   01-DEC-2000  (first entry)
XX
DE   Human androgen shutoff gene 3 (AS3) protein sequence.
XX
KW   Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;
KW   chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;
KW   diagnosis; treatment; cytostatic; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          419..459
FT                   /label= "Subdomain I"
FT                   /note= "Hank's conserved region"
FT   Region          419..425
FT                   /label= "Beta strand 1"
FT   Region          426..433
FT                   /label= "Mg-ATP binding loop"
FT   Region          453..459
FT                   /label= "Beta strand 2"
FT   Region          472..482
FT                   /label= "Subdomain II"
FT                   /note= "Hank's conserved domain - beta strand 3"
FT   Region          489..498
FT                   /label= "Subdomain III"
FT                   /note= "Hank's conserved region - alpha helix C"
```

```
FT   Region          509..516
FT                   /label= "Subdomain IV"
FT                   /note= "Hank's conserved region - beta strand 4"
FT   Region          525..548
FT                   /label= "Subdomain V"
FT                   /note= "Hank's conserved region"
FT   Region          525..532
FT                   /label= "Beta strand 5"
FT   Region          540..548
FT                   /label= "Al[ha helix D"
FT   Region          554..567
FT                   /label= "Subdomain VIa"
FT                   /note= "Hank's conserved region - alpha helix E"
XX
PN   WO200050454-A1.
XX
PD   31-AUG-2000.
XX
PF   24-FEB-2000; 2000WO-US04732.
XX
PR   24-FEB-1999;   99US-0121461.
XX
PA   (TUFT ) TUFTS COLLEGE.
XX
PI   Soto AM,  Sonnenschein C,  Geck P,  Szelei J;
XX
DR   WPI; 2000-565451/52.
DR   N-PSDB; AAA28051, AAA28052.
XX
PT   New human androgen-induced tumor suppressor cDNA sequence termed
PT   'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
PT   diagnosis and treatment of prostate cancer -
XX
PS   Claim 2; Fig 1; 152pp; English.
XX
CC   This invention relates to a human androgen-induced tumour suppressor cDNA
CC   sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is located
CC   on chromosome 13 at position 13q12-13q. AS3 has a role in inhibiting cell
CC   proliferation and use as a marker for the efficient diagnosis and
CC   treatment of prostate cancer. The invention includes AS3 cDNA and protein
CC   sequences, a vector comprising the cDNA sequence, a host cell transfected
CC   with the expression vector, and a method for producing an AS3 polypeptide
CC   comprising culturing the transfected cells. AS3 has cytostatic activity,
CC   and acts to suppress cell proliferation. The AS3 gene is useful as a
CC   marker for the efficient diagnosis and treatment of prostate cancer. The
CC   AS3 nucleic acid molecule can be used as a source of antisense agents for
CC   sequence specific modulation of gene expression. The AS3 protein may be
CC   used in the treatment of disorders caused by aberrant modification or
CC   mutation of a gene encoding an AS3 protein, misregulation of the AS3 gene
CC   or aberrant post-translational modification of the AS3 protein. This
CC   sequence represents the human AS3 protein sequence.
XX
SQ   Sequence   1391 AA;

  Query Match             100.0%;  Score 7193;  DB 21;  Length 1391;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1391;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLA 60

Qy         61 LHLASDFFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHLASDFFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDT 120

Qy        121 KSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDL 180

Qy        181 MSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTFFNQV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 MSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTFFNQV 240

Qy        241 LMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMF 300

Qy        301 GAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSH 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSH 360

Qy        361 DPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 DPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYA 420

Qy        421 LQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYL 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYL 480

Qy        481 YATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITRNLPDPG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 YATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVITRNLPDPG 540

Qy        541 KAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEM 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 KAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEM 600

Qy        601 IKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTH 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 IKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTH 660

Qy        661 PISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKK 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 PISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKK 720

Qy        721 GPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQ 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQ 780

Qy        781 FAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMK 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 FAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMK 840

Qy        841 NNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIIT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 NNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIIT 900

Qy        901 LEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQ 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 LEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQ 960

Qy        961 CLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKEC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKEC 1020

Qy       1021 LWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKS 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 LWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKS 1080

Qy       1081 TTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSS 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSS 1140

Qy       1141 AGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDK 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDK 1200

Qy       1201 RDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESD 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 RDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESD 1260

Qy       1261 EQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 EQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSG 1320
```

```
Qy       1321 PPAPEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESPESSAIESTQSTPQKGRGRPSK 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 PPAPEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESPESSAIESTQSTPQKGRGRPSK 1380

Qy       1381 TPSPSQPKKNV 1391
              |||||||||||
Db       1381 TPSPSQPKKNV 1391


RESULT 2
ABG72803
ID   ABG72803 standard; Protein; 1297 AA.
XX
AC   ABG72803;
XX
DT   24-FEB-2003  (first entry)
XX
DE   Human tumour supressor SCC-112.
XX
KW   Human; SCC-112; cancer; apoptosis; allodynia; cell proliferation;
KW   degenerative disorder; metastasis inhibition; breast cancer; causalgia;
KW   kidney cancer; bladder cancer; pancreatic cancer; colon cancer;
KW   squamous cell carcinoma; head trauma; spinal cord injury; herpes zoster;
KW   global and focal ischaemic and haemorrhagic stroke; epilepsy; neuralgia;
KW   hypoxia-induced nerve cell damage; anxiety; diabetes mellitus;
KW   cardiac arrest; spinal cord lesion; lung cancer; tumour supressor;
KW   neonatal distress; Alzheimer's disease; colon cancer; stomach cancer;
KW   multiple sclerosis; phantom limb pain; hyperalgesia; Down's syndrome;
KW   Huntington's disease; Parkinson's disease; Korsakoff's syndrome;
KW   amyotrophic lateral sclerosis; cell survival; uterine cancer.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          2..137
FT                   /label= RhoGEF domain
FT   Domain          127..651
FT                   /label= N-adaptin_domain
FT   Region          166..187
FT                   /label= Leucine_Zipper
FT   Domain          249..350
FT                   /label= SKP1_domain
FT   Region          597..617
FT                   /note= "PEST sequence site"
FT   Region          858..865
FT                   /note= "Tyrosine kinase phosphorylation site"
FT   Region          920..926
FT                   /note= "Nuclear localisation signature sequence"
FT   Region          1030..1036
FT                   /note= "Tyrosine kinase phosphorylation site"
FT   Region          1143..1663
FT                   /note= "PEST sequence site"
FT   Region          1216..1277
FT                   /note= "PEST sequence site"
FT   Region          1225..1231
FT                   /note= "Nuclear localisation signature sequence"
FT   Region          1227..1230
FT                   /note= "Nuclear localisation signature sequence"
FT   Region          1228..1234
FT                   /note= "Nuclear localisation signature sequence"
FT   Region          1232..1235
FT                   /note= "Nuclear localisation signature sequence"
FT   Region          1251..1257
FT                   /note= "Nuclear localisation signature sequence"
FT   Region          1278..1297
FT                   /label= Epitope_region
FT                   /note= "Residues 1278 to 1297 are
FT                           specifically claimed in claim 41"
XX
PN   WO200281641-A2.
XX
PD   17-OCT-2002.
XX
PF   08-APR-2002; 2002WO-US10850.
XX
PR   06-APR-2001; 2001US-281780P.
XX
PA   (GEOU ) UNIV GEORGETOWN.
XX
PI   Kasid UN,  Kumar D,  Ahmad I;
XX
DR   WPI; 2003-103330/09.
DR   N-PSDB; ABX14056.
XX
PT   New isolated SSC (undefined) tumor suppressor polypeptides and
PT   polynucleotides, useful for diagnosing, preventing or treating cancer
PT   or degenerative disease, e.g. Alzheimer's Disease, Huntington's
PT   disease, or multiple sclerosis  -
XX
PS   Claim 13; Fig 1B; 83pp; English.
XX
CC   The invention relates to a new isolated polypeptide SCC-112. The SCC-112
CC   polypeptides and polynucleotides are useful for diagnosing, preventing or
CC   treating cancer (e.g. breast cancer, kidney cancer, bladder cancer,
CC   pancreatic cancer, colon cancer, squamous cell carcinoma, uterine
CC   cancer, stomach cancer, colon cancer, lung cancer); or degenerative
CC   disease or disorder (e.g. global and focal ischaemic and haemorrhagic
CC   stroke, head trauma, spinal cord injury, hypoxia-induced nerve cell
CC   damage, nerve cell damage caused by cardiac arrest or neonatal distress,
CC   epilepsy, anxiety, diabetes mellitus, multiple sclerosis, phantom limb
CC   pain, causalgia, neuralgias, herpes zoster, spinal cord lesions,
CC   hyperalgesia, allodynia, Alzheimer's disease, Huntington's disease,
CC   Parkinson's disease, amyotrophic lateral sclerosis, Down's syndrome and
CC   Korsakoff's syndrome). The polypeptides and polynucleotides are also
CC   useful for inducing apoptosis in cancer cells, increasing survival or
CC   proliferation of a cell, or inhibiting cancer cell proliferation and/or
CC   metastasis in a cancer patient. The polynucleotides can be used as probes
CC   to detect complementary nucleotide sequences, or as primers to obtain
CC   additional copies of the polynucleotides. SSC-112 may also be used for
CC   identifying drugs for treatment of cancers. The present sequence
CC   represents the amino acid sequence of human tumour supressor SCC-112.
XX
SQ   Sequence   1297 AA;

  Query Match             64.5%;  Score 4642.5;  DB 24;  Length 1297;
  Best Local Similarity   68.8%;  Pred. No. 3.7e-299;
  Matches  905;  Conservative  159;  Mismatches  203;  Indels   49;  Gaps   11;

Qy         31 MVRRLKMVVKTFMDMDQDSEEEKELYLNLALHLASDFFLKHPGKDVRLLVACCLADIFRI 90
              |::|||||||||||||||||:||: || |||||||:|||::| |||||||||||||||||
Db          1 MIKRLKMVVKTFMDMDQDSEDEKQQYLPLALHLASEFFLRNPNKDVRLLVACCLADIFRI 60

Qy         91 YAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSYNICFELEDS 150
              ||||||||| |||||||:||||||||||||||||||||||||||:|||||||||||||| 
Db         61 YAPEAPYTSHDKLKDIFLFITRQLKGLEDTKSPQFNRYFYLLENLAWVKSYNICFELEDC 120

Qy        151 NEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSSIICEGDTVSQELLDTVLVNLVPAHKNL 210
              |||| ||:||||||||| ||:|| |||:||||||| ||  |:|||| ::|:||:||||||
Db        121 NEIFIQLFRTLFSVINNSHNKKVQMHMLDLMSSIIMEGCGVTQELLGSILINLIPAHKNL 180

Qy        211 NKQAYDLAKALLKRTAQAIEPYITTFFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLL 270
              |||::|||| ||||| | ||  |  ||||||:||::|:||||| ||||| ||: || |||
Db        181 NKQSFDLAKVLLKRTVQTIEACIANFFNQVLVLGRSSVSDLSEGVFDLIQELFAIDPHLL 240

Qy        271 LSVLPQLEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIR 330
              |||:||||||||||| |||| ||:||||:||:|||:||:||:|||||:||||||||||:|
Db        241 LSVMPQLEFKLKSNDGEERLAVVRLLAKLFGSKDSDLATQNRPLWQCFLGRFNDIHVPVR 300

Qy        331 LECVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLL 390
              || |||||||||||||||||||||||||||||||||||||||:|:||||:|: |||| ||
Db        301 LESVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVTIITAAKRDLALVNDQLL 360

Qy        391 NFVRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSID 450
               |||||||||||||||||||||||:|||| |   |||:||::::||||||||||||||||
Db        361 GFVRERTLDKRWRVRKEAMMGLAQLYKKYCLHGEAGKEAAEKVSWIKDKLLHIYYQNSID 420
```

```
Qy        451 DRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLL 510
              |:||||:|||||:|||||| |||||||||||:|| ||||||||||||||:||  |::||
Db        421 DKLLVEKIFAQYLVPHNLETEERMKCLYYLYASLDPNAVKALNEMWKCQNMLRSHVRELL 480

Qy        511 DLIKQPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPT 570
              || ||| ::|:  |:| |:| | :|||||||||||:||| ||| ||||:| |||:|:|||
Db        481 DLHKQPTSEANCSAMFGKLMTIAKNLPDPGKAQDFVKKFNQVLGDDEKLRSQLELLISPT 540

Qy        571 CSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQVNKSID 630
              ||||||: ||||| :|| |||||||||||:|||||||||||||:|:||||:| :||||:
Db        541 CSCKQADICVREIARKLANPKQPTNPFLEMVKFLLERIAPVHIDSEAISALVKLMNKSIE 600

Qy        631 GTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAALQ 690
              ||||||:|||  | |||:|||||||||||| |||||||:|||| ||:|:|:||||||:|
Db        601 GTADDEEEGVSPDTAIRSGLELLKVLSFTHPTSFHSAETYESLLQCLRMEDDKVAEAAIQ 660

Qy        691 IFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEP 750
              ||:||| ||| | | ||| |:|:|| |:|:| | ||| |:||||||||::|| | ||||||
Db        661 IFRNTGHKIETDLPQIRSTLIPILHQKAKRGTPHQAKQAVHCIHAIFTNKEVQLAQIFEP 720

Qy        751 LHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPWKSWVATFIVKDLLMNDRLPGKKTTK 810
              | :||:       | |||||::|||::|||||||:| || || |||||||||| |:| |
Db        721 LSRSLNADVPEQLITPLVSLGHISMLAPDQFASPMKSVVANFIVKDLLMNDRSTGEKNGK 780

Qy        811 LWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKI 870
              || |||||||| : |:||||::||||||||| |||  ||||||: :| |:||||| :|
Db        781 LWSPDEEVSPEVLAKVQAIKLLVRWLLGMKNNQSKSANSTLRLLSAMLVSEGDLTEQKRI 840

Qy        871 SKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGL 930
              || ||||||||||||:|||||||||||| ||:||||| |||||||||:||||||| |
Db        841 SKSDMSRLRLAAGSAIMKLAQEPCYHEIITPEQFQLCALVINDECYQVRQIFAQKLHKAL 900

Qy        931 SRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEY 990
               :| ||||||| ||||||||||||||||||||:|||::||||:||:    :|||||||||
Db        901 VKLLLPLEYMAIFALCAKDPVKERRAHARQCLLKNISIRREYIKQNPMATEKLLSLLPEY 960

Qy        991 VVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQT 1050
              |||| |||||||||: : ||::||:||||||:||:|| |||||||||||||::|| |||| |
Db        961 VVPYMIHLLAHDPDFTRSQDVDQLRDIKECLWFMLEVLMTKNENNSHAFMKKMAENIKLT 1020

Qy       1051 KDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKN 1110
              :||| ||::| |||||||||||: :| |||    : :||||||| :||||:  | | |:
Db       1021 RDAQSPDESKTNEKLYTVCDVALCVINSKSALCNADSPKDPVLPMKFFTQPEDKFCNDKS 1080

Qy       1111 YLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVS--NASSSSNPSSPG 1168
              |:  | :       ||||   ||||||||||: |::    :|:  || |  | :| |||:
Db       1081 YISEETRVLLLTGKPKPAGVLGAVNKPLSATGRKPYVRSTGTETGSNINVNSELNPSTGN 1140

Qy       1169 RIKGR-LDSSEMDHSENED--YTMSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQ 1225
              | :  :::|    ||||:    |  |  |  |                    ||    :
Db       1141 RSREQSSEAAETGVSENEENPVRIISVTPVKNID--------------------PVKNK 1179

Qy       1226 EEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESDE-QQWPEEKRLKEDILENEDEQNSP 1284
              |  :  |  |      |    | | :||  ||: ::  ||  :||          ||  |
Db       1180 E--INSDQAT-----QGNISSDRGKKRTVTAAGAENIQQKTDEK---------VDESGPP 1223

Qy       1285 -PKKGKRGRPPKPLG-GGTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEERQSGNTE 1338
                | | :||| ||    |    |  :    |   || |:      | :|      ::|| :
Db       1224 APSKPRRGRRPKSESQGNATKNDDLNKPINKGRKR----AAVGQESPGGLEAGNAK 1275
```

RESULT 3
ABP64685
ID   ABP64685 standard; Protein; 919 AA.
XX
AC   ABP64685;
XX
DT   25-FEB-2003 (first entry)
XX
DE   Human protein SEQ ID 345.
XX
KW   Human; expressed sequence tag; EST;
KW   haematopoietic disorder; central nervous system disease; viral infection;
KW   peripheral nervous system disease; non-healing wound; infectious disease;
KW   immune deficiency; immune disorder; bacterial infection; allergy; cancer;
KW   fungal infection; autoimmune disorder; coagulation disorder; nootropic;
KW   antiallergic; antiinflammatory; immunosuppressive; neuroprotective;
KW   cytostatic; haemostatic; virucide; antibacterial; fungicide;
KW   immunostimulant; cerebroprotective.
XX
OS   Homo sapiens.
XX
PN   WO200259260-A2.
XX
PD   01-AUG-2002.
XX
PF   16-NOV-2001; 2001WO-US42950.
XX
PR   17-NOV-2000; 2000US-0714936.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;
PI   Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR   WPI; 2002-590824/63.
DR   N-PSDB; ABQ99271.
XX
PT   New isolated polynucleotide, useful in research, diagnostic or
PT   therapeutic methods, e.g. preventing or treating disorders involving
PT   aberrant protein expression or biological activity -
XX
PS   Claim 20; SEQ ID 345; 394pp; English.
XX
CC   The present invention relates to novel human coding sequences
CC   (ABQ99268-ABQ99608) and proteins (ABP64682-ABP65022). The sequences are
CC   useful in therapeutic, diagnostic and research methods. The
CC   polynucleotides may be used in the field of molecular biology as
CC   hybridisation probes, primers for PCR, for chromosome and gene mapping,
CC   for the recombinant production of protein, or in generation of anti-sense
CC   DNA or RNA. The polynucleotides are useful in diagnostics as expressed
CC   sequence tags (ESTs) for identifying expressed genes or for physical
CC   mapping of the human genome. The proteins may be used as molecular weight
CC   markers, or as nutritional sources or supplements. The proteins may be
CC   used to maintain and expand cell population in a totipotential or
CC   pluripotential state useful for re-engineering damaged or diseased
CC   tissues, transplantation, manufacture of bio-pharmaceuticals or the
CC   development of bio-sensors. The polynucleotides and proteins are useful
CC   for preventing, treating or ameliorating disorders involving aberrant
CC   protein expression or biological activity, e.g. haematopoietic disorders,
CC   central/peripheral nervous system diseases, mechanical and traumatic
CC   disorders, non-healing wounds, immune deficiencies and disorders,
CC   infectious diseases caused by viral, bacterial or fungal infection,
CC   autoimmune disorders, allergic reactions and conditions, coagulation
CC   disorders, or cancer. The polynucleotide sequences of the invention were
CC   assembled from ESTs isolated mainly by sequencing by hybridisation, and
CC   in some cases, sequences obtained from one or more public databases.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   919 AA;

Query Match 41.7%; Score 2999.5; DB 23; Length 919;
Best Local Similarity 63.3%; Pred. No. 2.7e-190;
Matches 594; Conservative 123; Mismatches 172; Indels 49; Gaps 11;

```
Qy        409 MMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNL 468
              ||||||:|||| |    |||:||::::||||||||||||||||:||||:|||||:|||||
Db          1 MMGLAQLYKKYCLHGEAGKEAAEKVSWIKDKLLHIYYQNSIDDKLLVEKIFAQYLVPHNL 60

Qy        469 ETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSK 528
              || |||||||||||:|| ||||||||||||||:||  |::|||| ||| ::|:  |:| |
Db         61 ETEERMKCLYYLYASLDPNAVKALNEMWKCQNMLRSHVRELLDLHKQPTSEANCSAMFGK 120

Qy        529 VMVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLG 588
```

```
                 :| | :||||||||||:||| ||| ||||:| |||:|:|||||||||: ||||| :||
Db        121 LMTIAKNLPDPGKAQDFVKKFNQVLGDDEKLRSQLELLISPTCSCKQADICVREIARKLA 180

Qy        589 NPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRA 648
              ||||||||||||:||||||||||||||:|:||||:| :||||:|||||||:|||  | |||:
Db        181 NPKQPTNPFLEMVKFLLERIAPVHIDSEAISALVKLMNKSIEGTADDEEEGVSPDTAIRS 240

Qy        649 GLELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRS 708
              ||||||||||||| ||||||:|||| ||:|:|:||||||:|||:||| ||| | | |||
Db        241 GLELLKVLSFTHPTSFHSAETYESLLQCLRMEDDKVAEAAIQIFRNTGHKIETDLPQIRS 300

Qy        709 ALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLV 768
               |:|:|| |:|:| | ||| |:|||||||::|| | ||||||| :||:     | ||||||
Db        301 TLIPILHQKAKRGTPHQAKQAVHCIHAIFTNKEVQLAQIFEPLSRSLNADVPEQLITPLV 360

Qy        769 TIGHIALLAPDQFAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQA 828
              ::|||::|||||||:| || || ||||||||||   |:|  ||| |||||||| : |:||
Db        361 SLGHISMLAPDQFASPMKSVVANFIVKDLLMNDRSTGEKNGKLWSPDEEVSPEVLAKVQA 420

Qy        829 IKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVK 888
              ||::||||||||||| |||  ||||||: :| |:|||||| :||| |||||||||||||:|
Db        421 IKLLVRWLLGMKNNQSKSANSTLRLLSAMLVSEGDLTEQKRISKSDMSRLRLAAGSAIMK 480

Qy        889 LAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAK 948
              ||||||||||| ||:||||| ||||||||||:||||||| | :| |||||||| |||||
Db        481 LAQEPCYHEIITPEQFQLCALVINDECYQVRQIFAQKLHKALVKLLLPLEYMAIFALCAK 540

Qy        949 DPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKV 1008
              ||||||||||||||:|||::||||:||:    :||||||||||||||| ||||||||||: :
Db        541 DPVKERRAHARQCLLKNISIRREYIKQNPMATEKLLSLLPEYVVPYMIHLLAHDPDFTRS 600

Qy       1009 QDIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTV 1068
              ||::||:|:||||||:||:|| |||||||||::|| |||| |:||| ||::| |||||||
Db        601 QDVDQLRDIKECLWFMLEVLMTKNENNSHAFMKKMAENIKLTRDAQSPDESKTNEKLYTV 660

Qy       1069 CDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTT 1128
              ||||: :| |||    : :|||||||| :||||:|:| | |:|:  | :       ||||
Db        661 CDVALCVINSKSALCNADSPKDPVLPMKFFTQPEKDFCNDKSYISEETRVLLLTGKPKPA 720

Qy       1129 NVLGAVNKPLSSAGKQSQTKSSRMETVS--NASSSSNPSSPGRIKGR-LDSSEMDHSENE 1185
               ||||||||||: |::    :|:  || |  | :|  |||:  | : :  :::|   ||||
Db        721 GVLGAVNKPLSATGRKPYVRSTGTETGSNINVNSELNPSTGNRSREQSSEAAETGVSENE 780

Qy       1186 D--YTMSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKP 1243
              :     : | | | |                          || :|  : | |      |
Db        781 ENPVRIISVTPVKNID--------------------PVKNKE--INSDQAT-----QGN 812

Qy       1244 KGSQRSRKRGHTASESDE-QQWPEEKRLKEDILENEDEQNSP-PKKGKRGRPPKPLG-GG 1300
               | | :|| ||: :: || :||            ||  | | | :||| ||   |
Db        813 ISSDRGKKRTVTAAGAENIQQKTDEK---------VDESGPPAPSKPRRGRRPKSESQGN 863

Qy       1301 TPKEEPTMKTSKKGSKKKSGPPAPEEEEEEERQSGNTE 1338
               | :  |  ||| |:     | :|     ::|| :
Db        864 ATKNDDLNKPINKGRKR----AAVGQESPGGLEAGNAK 897
```

RESULT 4
AAB01382
ID   AAB01382 standard; Protein; 854 AA.
XX
AC   AAB01382;
XX
DT   20-OCT-2000 (first entry)
XX
DE   Neuron-associated protein.
XX
KW   Neuron associated protein; NEUAP; neurological disorder; epilepsy;
KW   ischemic cerebrovascular disease; stroke; cerebral neoplasm;
KW   Alzheimer's disease; Pick's disease; Huntington's disease;
KW   dementia; Parkinson's disease; demyelinating disease; meningitis;
KW   prion disease; kuru; Creutzfeldt-Jakob disease; neurofibromatosis;
KW   cerebral palsy; muscular dystrophy; central nervous system; CNS;
KW   peripheral nervous system; PNS; myopathy; schizophrenia;
KW   actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;
KW   cirrhosis; hepatitis; mixed connective tissue disease; MCTD;
KW   myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;
KW   autoimmune disease; inflammation; acquired immunodeficiency syndrome;
KW   AIDS; Addison's disease; adult respiratory distress syndrome;
KW   allergy; ankylosing spondylitis; amyloidosis; anaemia; asthma;
KW   Werner syndrome, trauma; human.
XX
OS   Homo sapiens.
XX

```
FH   Key             Location/Qualifiers
FT   Modified-site   48
FT                   /note= "Potential glycosylation site"
FT   Modified-site   88
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   99
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   153
FT                   /note= "Potential glycosylation site"
FT   Modified-site   159
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   175
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   249
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   265
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   330
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   340
FT                   /note= "Potential phosphorylation site"
FT   Region          358..409
FT                   /label= "Postsynaptic protein signature"
FT   Modified-site   369
FT                   /note= "Potential glycosylation site"
FT   Region          373..422
FT                   /label= "Neuraxin signature"
FT   Modified-site   375
FT                   /note= "Potential glycosylation site"
FT   Modified-site   377
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   387
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   398
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   422
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   492
FT                   /note= "Potential glycosylation site"
FT   Modified-site   494
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   507
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   557
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   561
FT                   /note= "Potential glycosylation site"
FT   Modified-site   582
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   593
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   594
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   614
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   626
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   649
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   668
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   677
FT                   /note= "Potential phosphorylation site"
```

```
FT   Modified-site   697
FT                   /note= "Potential glycosylation site"
FT   Modified-site   712
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   747
FT                   /note= "Potential glycosylation site"
FT   Modified-site   750
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   757
FT                   /note= "Potential phosphorylation site"
FT   Modified-site   798
FT                   /note= "Potential glycosylation site"
FT   Modified-site   800
FT                   /note= "Potential phosphorylation site"
XX
PN   WO200034477-A2.
XX
PD   15-JUN-2000.
XX
PF   10-DEC-1999;   99WO-US30408.
XX
PR   11-DEC-1998;   98US-0210083.
PR   11-DEC-1998;   98US-9123456.
PR   09-FEB-1999;   99US-0119365.
PR   16-MAR-1999;   99US-0124687.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Tang YT,  Yue H,  Baughn MR,  Hillman JL,  Lal P,  Au-young J,  Yang J;
PI   Lu DAM,  Azimzai Y;
XX
DR   WPI; 2000-423423/36.
DR   N-PSDB; AAA47423.
XX
PT   New human neuron-associated proteins and polynucleotides encoding them,
PT   useful for diagnosis, treatment and prevention of cell proliferative
PT   disorders including cancer, neuronal and neurological disorders
XX
PS   Claim 1; Page 103-105; 145pp; English.
XX
CC   Human neuron-associated proteins (NEUAP) can be used for for
CC   treating or preventing a disorder associated with decreased
CC   expression or activity of NEUAP. Antagonists of NEUAP are useful for
CC   treating or preventing disorder associated with increased expression
CC   or activity of NEUAP. NEUAP or their fragments or derivatives are
CC   useful for treating neurological disorder such as epilepsy, ischemic
CC   cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's
CC   disease, Pick's disease, Huntington's disease, dementia and
CC   Parkinson's disease. NEUAPs are also useful for treating other
CC   demyelinating diseases, bacterial and viral meningitis, prion
CC   diseases including kuru, Creutzfeldt-Jakob disease, nutritional and
CC   metabolic diseases of the nervous system, neurofibromatosis, other
CC   developmental disorders of the central nervous system, cerebral
CC   palsy, neuroskeletal disorders, autonomic nervous system disorders,
CC   cranial nerve disorders, spinal cord diseases, muscular dystrophy and
CC   other neuromuscular disorders, peripheral nervous system disorders,
CC   inherited, metabolic, endocrine, and toxic myopathies, mental
CC   disorders including mood, anxiety and schizophrenic disorders, a cell
CC   proliferative disorder such as actinic keratosis, arteriosclerosis,
CC   atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective
CC   tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal
CC   haemoglobinuria, cancers of the adrenal gland, bladder, bone,
CC   bone marrow, brain, breast, cervix, and an autoimmune/inflammatory
CC   disorder such as acquired immunodeficiency syndrome (AIDS), Addison's
CC   disease, adult respiratory distress syndrome, allergies, ankylosing
CC   spondylitis, amyloidosis, anemia, asthma, Werner syndrome,
CC   complications of cancer, hemodialysis, and extracorporeal circulation,
CC   viral, bacterial, fungal parasitic, protozoal, and helminthic
CC   infections, and trauma. This protein was given the Incyte ID no.
CC   2888437CD1.
XX
SQ   Sequence   854 AA;

  Query Match             37.8%;  Score 2715.5;  DB 21;  Length 854;
  Best Local Similarity   62.2%;  Pred. No. 1.7e-171;
  Matches  543;  Conservative  114;  Mismatches  167;  Indels   49;  Gaps   11;

Qy        474 MKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMVIT 533
Db          1 MKCLYYLYASLDPNAVKALNEMWKCQNMLRIHVRELLDLHKQPTSEANCSAMFGKLMTIA 60

Qy        534 RNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQP 593
Db         61 KNLPDPGKAQDFVKKFNQVLGDDEKLRSQLELLISPTCSCKQADICVREIARKLANPKQP 120

Qy        594 TNPFLEMIKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELL 653
Db        121 TNPFLEMVKFLLERIAPVHIDSEAISALVKLMNKSIEGTADDEEEGVSPDTAIRSGLELL 180

Qy        654 KVLSFTHPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPV 713
Db        181 KVLSFTHPTSFHSAETYESLLQCLRMEDDKVAEAAIQIFRNTGHKIETDLPQIRSTLIPI 240

Qy        714 LHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHI 773
Db        241 LHQKAKRGTPHQAKQAVHCIHAIFTNKEVQLAQIFEPLSRSLNADVPEQLITPLVSLGHI 300

Qy        774 ALLAPDQFAAPWKSWVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMV 833
Db        301 SMLAPDQFASPMKSVVANFIVKDLLMNDRSTGEKNGKLWSPDEEVSPEVLAKVQAIKLLV 360

Qy        834 RWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEP 893
Db        361 RWLLGMKNNQSKSANSTLRLLSAMLVSEGDLTEQKRISKSDMSRLRLAAGSAIMKLAQEP 420

Qy        894 CYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKE 953
Db        421 CYHEIITPEQFQLCALVINDECYQVRQIFAQKLHKALVKLLLPLEYMAIFALCAKDPVKE 480

Qy        954 RRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQ 1013
Db        481 RRAHARQCLLKNISIRREYIKQNPMATEKLLSLLPEYVVPYMIHLLAHDPDFTRSQDVDQ 540

Qy       1014 LKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAM 1073
Db        541 LRDIKECLWFMLEVLMTKNENNSHAFMKKMAENIKLTRDAQSPDESKTNEKLYTVCDVAL 600

Qy       1074 NIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGA 1133
Db        601 CVINSKSALCNADSPKDPVLPMKFFTQPEKDFCNDKSYISEETRVLLLTGKPKPAGVLGA 660

Qy       1134 VNKPLSSAGKQSQTKSSRMETVS--NASSSSNPSSPGRIKGR-LDSSEMDHSENED--YT 1188
Db        661 VNKPLSATGRKPYVRSTGTETGSNINVNSELNPSTGNRSREQSSEAAETGVSENEENPVR 720

Qy       1189 MSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQR 1248
Db        721 IISVTPVKNID--------------------PVKNKE--INSDQAT-----QGNISSDR 752

Qy       1249 SRKRGHTASESDE-QQWPEEKRLKEDILENEDEQNSP-PKKGKRGRPPKPLG-GGTPKEE 1305
Db        753 GKKRTVTAAGAENIQQKTDEK---------VDESGPPAPSKPRRGRPKSESQGNATKND 803

Qy       1306 PTMKTSKKGSKKKSGPPAPEEEEEEERQSGNTE 1338
Db        804 DLNKPINKGRKR----AAVGQESPGGLEAGNAK 832


RESULT 5
ABB67869
ID   ABB67869 standard; Protein; 1218 AA.
XX
AC   ABB67869;
XX
DT   26-MAR-2002  (first entry)
XX
```

```
DE   Drosophila melanogaster polypeptide SEQ ID NO 30399.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL11972.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 30399; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1218 AA;

  Query Match             27.6%;  Score 1982.5;  DB 22;  Length 1218;
  Best Local Similarity   37.3%;  Pred. No. 1.3e-122;
  Matches  444;  Conservative  247;  Mismatches  421;  Indels   77;  Gaps   22;

Qy         13 ITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLALHLASDFFLKHP 72
              | || |  : : : :  :|::||||  :            ||||     :: |: ||||| |||::||
Db          4 IVYPTGCRPLVEDLGTDELIRRLKTLANVLQTMDQDDNLYQQ-YIPLALHLLDDFFMQHP 62

Qy         73 GKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQFNRYFYLL 132
               :||:||:|||:||: |:|||||||   |::| || |   :|| ||:| : | | ||||||
Db         63 SRDVQLLIACCVADVLRVYAPEAPYKEQDQIKTIFKFFIKQLHGLKDPRDPSFKRYFYLL 122

Qy        133 ENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMSSIICEGDTVS 192
              ||:|:|||:|:||||||  |||  |: |:|  ::|: |: ||     :|::|  :| |  | :|
Db        123 ENLAFVKSFNMCFELEDCQEIFQDLFSTIFKIVNDQHSVKVTNFFLDVLSPLITEADNLS 182

Qy        193 QELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITTFFNQVLMLGK--TSISD 250
               ||||  :|:|:|   :|: || |    |  :  || :|   |:|     |  |||: |:: |  | :|
Db        183 VELLDLILINIVEPYKSNNKFACQLTEQLLTKTGDALESTIKMFFNRALVMDKPNTKLS- 241

Qy        251 LSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQ 310
              ::   ::|:| ||   |:: || |||||||  || |  |: |||:     ||::||  |||:|| :
Db        242 ITNKIYDIIYELNRINAGLLCSVLPQLENKLLSTDDAERLKATTLLSRMFSEKDSQLAKK 301

Qy        311 NKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDV 370
                 | : :  ||| ||   |:|::||: : |  |:|||  |   |:||  |::|:|| :| :||:|
Db        302 YPNLLKIFFGRFCDITEPVRIKCVQSSMHFLLNHPSLQHDITEKLRLRNHDLDEVVRHEV 361

Qy        371 IVSIVTAAKKDILLVND--HLLNFVRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKD 428
              :::||  ||:|  || :     ||  ||||||||::::|::|| ||| |||:    :
Db        362 VMAIVETAKRDFTLVLEAPDLLEIVRERTLDKKYKIRRDAMNGLAYIYKRAICEPNDLST 421

Qy        429 AAK-QIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLN 487
               | ::  |||:|:|| ||:   ::|||||||:     :||: |    |||| ||:|    || |
Db        422 GLKVRVDWIKNKILHGYYKVGLEDRLLVERLLITCLVPYKLAPEERMKKLYHLLGDLDAN 481

Qy        488 AVKALNEMWKCQNLLRHQVKDLLDLIKQPK-TDASVKAIFSKVMVITRNLPDPGKAQDFM 546
               | ||  |: | |   |: | | : |    : |    :   : :|   | : |||| || :::
Db        482 ATKAFVELQKNQMKTRNTVSDWIKLHHSKEFTPRVLSQLSAKQANIAKLLPDPLKAAEYL 541

Qy        547 KKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLLE 606
               :|:  |  | :: : : :::   ||::    :  : ||||   | :| :    :| |:|
Db        542 TQFSNNLRKDAQLLRCINIVLKRDVSCRECADTMGVLLKKLGAHVQ-SNLYYNTVKMLIE 600

Qy        607 RIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHS 666
              |:| | :| ||| ||  : : |:  :  |: |:     :|  ||:|| :||:      | :
Db        601 RVASVMVDKESIGVLISLIEQCIEKGSMCEEIGISAQEAGERGLKLLAMLSYVFSAHFFT 660

Qy        667 AETFESLLACLKMDDEKVAEAALQIFKNTG---SKIEEDFPHIRSALLPVLHHKSKKGPP 723
                :    |:: |  : : ||  |:  :  |     |::  |  |   | ||   :  | |
Db        661 DTSLRHLISLLSYEQDYVAPLVLKTLTHLGRYQPLIDDPTPAILDELAPVCKDFALIGTP 720

Qy        724 RQAKYAIHCIH--------------AIFSSKETQ-----FAQIFEPLHKSLDPSNLEHL 763
               :|||:|: ||                 |  :|  ||     |  :| | |   | | | | ||
Db        721 KQAKHAVRCIFVNSQSSASTDGATSGAGSASTTTQTVHPIFNEIIETLRLKLTP-NCEHQ 779

Qy        764 ITPLVTIGHIALLAPDQFAAPWKSWVATFIVKDLLMNDRLPGKKTTKL-----WVPDEEV 818
                | :||:||||   |   |  | |: :|  |||:||: : :| ::  :|        |   |::
Db        780 RTKIVTLGHIAYNMPQAFLTPIKNMIARRIVKELLIQE-VPAQRDYELPEDSDWCAQEKL 838

Qy        819 SPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRL 878
               |:|: |: |:| | ||||:: :     |  | |:|      ::   |||   | ::   | : |
Db        839 PPDTLCKLDALKAMARWLLGLRTD-EHAAQKTFRMLAAFVNQRGDLLGQNRLCGAEKSWL 897

Qy        879 RLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLR---- 934
               || |   |::|: ::     :   : |||    :   |    :||::||:||||||||
Db        898 RLGAACAMLKVCEQKGVGDQYSAEQYLQLSQLMADPVPEVREIFARKLHKGLSRSLPRNC 957

Qy        935 LPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVS------EKLLSLLP 988
              |||::| :   |    :  :: : :   |        ::| ||||||   |  |    : | :||
Db        958 LPLDFMGLYVLAGLETERKLQDLVRHYAETDVNKRREYLKTVAMTSPDSSTESQSLHILP 1017

Qy        989 EYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIK 1048
              :|:: : | :| |||  :   :|  ||: :::|| |:|| ||||  |    |:| ::::: ||
Db       1018 DYMLAFAIPVLVHDPRFTNHEDYVQLRKMEKCLRFILEPLMAKRETFVHSFYKQLLQLIK 1077

Qy       1049 QTKDAQGPDDAKMNEKLYTVCDVAMNIIMSK-------STTYSLESPKDPVLPARFFTQP 1101
               : : |  |  :  | |:: ::||:|| || ||          ::|:|:          || :: :|
Db       1078 HREFSLG-SDKRDNYKMWALCDLAMYIIDSKFSPFDGNTSTFSMPL----ALPEMYYKEP 1132

Qy       1102 D-KNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKS 1149
                   || |    |:| :: :                    |||  |  | |      :  |  |
Db       1133 AVANFQNNDVYIPLDVYT------------LGA--KSTSKAAATAMTTS 1167


RESULT 6
AAU27827
ID   AAU27827 standard; Protein; 217 AA.
XX
AC   AAU27827;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human full-length polypeptide sequence #152.
XX
KW   Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW   mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW   cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW   nervous system disorder; inflammatory disorder; cell differentiation;
KW   angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW   genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW   cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;
KW   antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
```

```
KW   neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;
KW   immunostimulant; analgesic; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200164834-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04926.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
PR   17-JUN-2000; 2000US-0597707.
PR   14-JUL-2000; 2000US-0616807.
PR   19-SEP-2000; 2000US-0664641.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Wehrman T,  Wang J,  Ma Y,  Wang D,  Chen R,  Xu C;
PI   Drmanac R;
XX
DR   WPI; 2001-589862/66.
DR   N-PSDB; AAS44727.
XX
PT   Novel polypeptides and nucleic acids obtained from cDNA libraries
PT   prepared from various human tissues, for diagnosis, treatment of
PT   cancer, neurological, inflammatory disorders and for use in arrays for
PT   detection  -
XX
PS   Claim 10; SEQ ID No 324; 153pp; English.
XX
CC   Sequences AAU27676-AAU28019 represent full-length polypeptides and
CC   contig polypeptides of the invention. The proteins and their associated
CC   DNA sequences are useful for the treatment, diagnosis and prevention of
CC   various types of disorder in a mammalian subject such as a human, dog,
CC   monkey, mouse, hamster or rat. The disorders include cancers such as
CC   leukaemia, lymphoma and neuroblastoma, autoimmune disorders such as
CC   multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC   diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC   disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC   chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC   Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC   disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC   bowel disease. The sequences exhibit activity relating to angiogenesis,
CC   cell proliferation, cell differentiation, stem cell growth factor,
CC   activin or inhibin. Therefore, they can be used to manipulate stem cells
CC   in culture to give rise to neuroepithelial cells that can be used to
CC   augment or replace cells damaged by illness, accidental damage or genetic
CC   disorders. The sequences may also be used for regeneration of bone,
CC   cartilage, tendons and ligaments and in tissue repair and burn healing.
CC   Note: Some sequences for this patent did not form part of the printed
CC   specification, but were obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   217 AA;

  Query Match             11.7%;  Score 839;  DB 22;  Length 217;
  Best Local Similarity   100.0%;  Pred. No. 9.7e-48;
  Matches  161;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1231 MDDLTKLVQEQKPKGSQRSRKRGHTASESDEQQWPEEKRLKEDILENEDEQNSPPKKGKR 1290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MDDLTKLVQEQKPKGSQRSRKRGHTASESDEQQWPEEKRLKEDILENEDEQNSPPKKGKR 60

Qy       1291 GRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEERQSGNTEQKSKSKQHRVSR 1350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEERQSGNTEQKSKSKQHRVSR 120

Qy       1351 RAQQRAESPESSAIESTQSTPQKGRGRPSKTPSPSQPKKNV 1391
              |||||||||||||||||||||||||||||||||||||||||
Db        121 RAQQRAESPESSAIESTQSTPQKGRGRPSKTPSPSQPKKNV 161
```

```
RESULT 7
AAB58462
ID   AAB58462 standard; Protein; 363 AA.
XX
AC   AAB58462;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Lung cancer associated polypeptide sequence SEQ ID 800.
XX
KW   Human; lung cancer associated protein; neuroprotective; cytostatic;
KW   cardioactive; immunomodulatory; muscular active; vulnerary;
KW   gastrointestinal; nephrotropic; antiinfective; gynecological;
KW   antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW   proliferative disorder; wound healing; infectious disease.
XX
OS   Homo sapiens.
XX
PN   WO200055180-A2.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05918.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Ruben SM;
XX
DR   WPI; 2000-587514/55.
DR   N-PSDB; AAF18338.
XX
PT   Lung cancer associated gene sequences, referred to as lung cancer
PT   antigens, useful for treatment, prevention, and diagnosis of disorders
PT   such as lung cancer -
XX
PS   Claim 11; Page 1337-1339; 1425pp; English.
XX
CC   Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC   associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC   associated proteins and polynucleotide sequences, their agonists, and
CC   antagonists may have neuroprotective; cytostatic; cardioactive;
CC   immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC   general; nephrotropic; antiinfective; gynecological; or antibacterial
CC   activity. The invention also includes antibodies specific for the
CC   protein or polynucleotide sequences. The lung cancer associated
CC   polynucleotide sequences may be used for detection of lung cancer,
CC   chromosome identification, as chromosome markers, and for numerous other
CC   diagnostic or research purposes. The proteins may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders. The proteins may also be used in the treatment of wounds and
CC   infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC   peptide AAB58549 are used in the course of the invention for the
CC   identification and characterisation of the polynucleotide and protein
CC   sequences.
XX
SQ   Sequence   363 AA;

  Query Match             10.5%;  Score 758.5;  DB 21;  Length 363;
  Best Local Similarity   45.0%;  Pred. No. 4.3e-42;
  Matches  172;  Conservative   56;  Mismatches  105;  Indels   49;  Gaps   11;

Qy        965 NINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFV 1024
              ||::||||:||:   :|||||||||||||| |||||||||: : ||::||:|:||||||:
Db          1 NISIRREYIKQNPMATEKLLSLLPEYVVPYMIHLLAHDPDFTRSQDVDQLRDIKECLWFM 60

Qy       1025 LEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYS 1084
              ||:|| |||||||||::|| |||| |:||| ||::| |||||||||||: :| |||   :
```

```
Db         61 LEVLMTKNENNSHAFMKKMAENIKLTRDAQSPDESKTNEKLYTVCDVALCVINSKSALCN 120

Qy       1085 LESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQ 1144
              :||||||| :|||||:|:| | |:|:  | :      ||||  |||||||||: |::
Db        121 ADSPKDPVLPMKFFTQPEKDFCNDKSYISEETRVLLLTGKPKPAGVLGAVNKPLSATGRK 180

Qy       1145 SQTKSSRMETVS--NASSSSNPSSPGRIKGR-LDSSEMDHSENED--YTMSSPLPGKKSD 1199
                 :|:  || |  | :|  |||:  | : :  :::|   ||||:    : |  | |  |
Db        181 PYVRSTGTETGSNINVNSELNPSTGNRSREQSSEAAETGVSENEENPVRIISVTPVKNID 240

Qy       1200 KRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASES 1259
                                  ||  :|  :  |  |     |     | | :||  ||: :
Db        241 --------------------PVKNKE--INSDQAT-----QGNISSDRGKKRTVTAAGA 272

Qy       1260 DE-QQWPEEKRLKEDILENEDEQNSP-PKKGKRGRPPKPLG-GGTPKEEPTMKTSKKGSK 1316
               :  ||  :||           ||   | | | :||| ||    |    | :   |  || |
Db        273 ENIQQKTDEK---------VDESGPPAPSKPRRGRRPKSESQGNATKNDDLNKPINKGRK 323

Qy       1317 KKSGPPAPEEEEEEERQSGNTE 1338
              :     |  :|     ::|| :
Db        324 R----AAVGQESPGGLEAGNAK 341
```

RESULT 8
ABB84560

```
ID   ABB84560 standard; Protein; 182 AA.
XX
AC   ABB84560;
XX
DT   23-JAN-2003  (first entry)
XX
DE   Human euchromosome fragile intelligence delay syndrome protein 20.02.
XX
KW   Euchromosome fragile intelligence delay syndrome protein 20.02; human;
KW   euchromosome fragile intelligence delay syndrome; dementia.
XX
OS   Homo sapiens.
XX
PN   CN1351041-A.
XX
PD   29-MAY-2002.
XX
PF   26-OCT-2000; 2000CN-0125797.
XX
PR   26-OCT-2000; 2000CN-0125797.
XX
PA   (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX
PI   Mao Y,  Xie Y;
XX
DR   WPI; 2002-637117/69.
DR   N-PSDB; ABS56684.
XX
PT   Human euchromosome fragile intelligence delay syndrome associated
PT   protein 20.02 polypeptide, used to treat e.g. dementia -
XX
PS   Claim 1; Page 30 (disclosure); 33pp; Chinese.
XX
CC   This invention describes the novel human euchromosome fragile
CC   intelligence delay syndrome associated protein, 20.02. The polypeptide is
CC   used in treating diseases such as euchromosome fragile intelligence delay
CC   syndrome, and dementia. This sequence represents the human euchromosome
CC   fragile intelligence delay syndrome-associated protein, 20.02 described
CC   in the disclosure of the invention.
XX
SQ   Sequence   182 AA;

  Query Match              9.9%;  Score 709;  DB 23;  Length 182;
  Best Local Similarity   75.0%;  Pred. No. 3.2e-39;
  Matches  129;  Conservative   24;  Mismatches   19;  Indels    0;  Gaps    0;

Qy        409 MMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNL 468
              ||||||:|||| |    |||:||::::||||||||||||||||:||||:|||||:|||||
Db          1 MMGLAQLYKKYCLHGEAGKEAAEKVSWIKDKLLHIYYQNSIDDKLLVEKIFAQYLVPHNL 60

Qy        469 ETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSK 528
              || |||||||||||:|| ||||||||||||||:||   |::||||  |||  ::|:  |:| |
Db         61 ETEERMKCLYYLYASLDPNAVKALNEMWKCQNMLRSHVRELLDLHKQPTSEANCSAMFGK 120

Qy        529 VMVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCV 580
              :| |  :||||||||||||:||| ||| ||||:| |||:|:|||||||||: ||
Db        121 LMTIAKNLPDPGKAQDFVKKFNQVLGDDEKLRSQLELLISPTCSCKQADICV 172
```

RESULT 9
AAB94946

```
ID   AAB94946 standard; Protein; 333 AA.
XX
AC   AAB94946;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:16468.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 16468; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
```

```
XX
SQ   Sequence   333 AA;

  Query Match             8.9%;  Score 640.5;  DB 22;  Length 333;
  Best Local Similarity   42.7%;  Pred. No. 2.6e-34;
  Matches  150;  Conservative   51;  Mismatches  101;  Indels   49;  Gaps   11;

Qy        996 IHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQG 1055
              |||||||||: : ||::||:|:||||||:||:|| |||||||||::|| |||| |:|||
Db          2 IHLLAHDPDFTRSQDVDQLRDIKECLWFMLEVLMTKNENNSHAFMKKMAENIKLTRDAQS 61

Qy       1056 PDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPE 1115
              ||::| |||||||||||: :| |||    : :|||||||| :|||||:|:| | |:|:  |
Db         62 PDESKTNEKLYTVCDVALCVINSKSALCNADSPKDPVLPMKFFTQPEKDFCNDKSYISEE 121

Qy       1116 MKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVS--NASSSSNPSSPGRIKGR 1173
                :       ||||    |||||||||: |::     :|:  || |    | :|  |||:  | : :
Db        122 TRVLLLTGKPKPAGVLGAVNKPLSATGRKPYVRSTGTETGSNINVNSELNPSTGNRSREQ 181

Qy       1174 -LDSSEMDHSENED--YTMSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKLG 1230
                :::|    ||||:      : |   | | |                     ||   :|   :
Db        182 SSEAAETGVSENEENPVRIISVTPVKNID--------------------PVKNKE--IN 218

Qy       1231 MDDLTKLVQEQKPKGSQRSRKRGHTASESDE-QQWPEEKRLKEDILENEDEQNSP-PKKG 1288
                |   |       |     | | :||  ||: ::   ||   :||             ||     | | |
Db        219 SDQAT-----QGNISSDRGKKRTVTAAGAENIQQKTDEK---------VDESGPPAPSKP 264

Qy       1289 KRGRPPKPLG-GGTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEERQSGNTE 1338
                :||| ||      |      |  :     |     || |:          |   :|       ::|| :
Db        265 RRGRRPKSESQGNATKNDDLNKPINKGRKR----AAVGQESPGGLEAGNAK 311
```

```
RESULT 10
AAU01759
ID   AAU01759 standard; Protein; 165 AA.
XX
AC   AAU01759;
XX
DT   18-JUL-2001  (first entry)
XX
DE   Human secreted protein #38.
XX
KW   Human; secreted protein; immunogen; antibody; diagnosis;
KW   rheumatoid arthritis; hyperproliferative disorder; neoplasm;
KW   cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW   angiogenesis; Alzheimer's disease; bacterial infection; viral infection;
KW   fungal infection; corneal infection; wound healing; cell culture;
KW   epithelial cell proliferation; skin ageing; transplantation;
KW   tissue regeneration; chemotaxis; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200123546-A1.
XX
PD   05-APR-2001.
XX
PF   26-SEP-2000; 2000WO-US26323.
XX
PR   27-SEP-1999;    99US-0155805.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben S,  Komatsoulis GA;
XX
DR   WPI; 2001-266150/27.
DR   N-PSDB; AAS02396.
XX
PT   Nucleic acids encoding 37 human secreted polypeptides, useful for
PT   preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT   disease and diabetic retinopathy -
XX
PS   Disclosure; Page 454; 494pp; English.
XX
```

```
CC   The sequence represents a human secreted protein of the invention. The
CC   polynucleotides, polypeptides and antibodies raised against them are used
CC   to prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. The
CC   polynucleotides and antibodies are also used in diagnosing a pathological
CC   condition or susceptibility to a pathological condition. The antibodies
CC   can also be used in alleviating symptoms associated with the disorders
CC   and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC   immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC   include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC   disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC   e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC   angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC   infections caused by bacteria, viruses and fungi and ocular disorders
CC   e.g. corneal infection. The polypeptides can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin aging due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC   polypeptides can also be used as a food additive or preservative to
CC   increase or decrease storage capabilities.
XX
SQ   Sequence   165 AA;

  Query Match             8.3%;  Score 599;  DB 22;  Length 165;
  Best Local Similarity   65.0%;  Pred. No. 5.7e-32;
  Matches  106;  Conservative   29;  Mismatches   28;  Indels    0;  Gaps    0;

Qy        964 KNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWF 1023
              :||::||||:||:    :||||||||||||| |||||||||: : ||::||:|:||||||
Db          1 ENISIRREYIKQNPMATEKLLSLLPEYVVPYMIHLLAHDPDFTRSQDVDQLRDIKECLWF 60

Qy       1024 VLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTY 1083
              :||:|| |||||||||::|| |||| |:||| ||::| |||||||||||: :| |||
Db         61 MLEVLMTKNENNSHAFMKKMAENIKLTRDAQSPDESKTNEKLYTVCDVALCVINSKSALC 120

Qy       1084 SLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPK 1126
              : :|||||||| :|||||:|:| | |:|:  | :       ||||
Db        121 NADSPKDPVLPMKFFTQPEKDFCNDKSYISEETRVLLLTGKPK 163
```

```
RESULT 11
ABB89260
ID   ABB89260 standard; Protein; 147 AA.
XX
AC   ABB89260;
XX
DT   24-MAY-2002  (first entry)
XX
DE   Human polypeptide SEQ ID NO 1636.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein.
XX
OS   Homo sapiens.
XX
PN   WO200190304-A2.
XX
PD   29-NOV-2001.
XX
PF   18-MAY-2001; 2001WO-US16450.
XX
PR   19-MAY-2000; 2000US-205515P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE,  Rosen CA;
XX
DR   WPI; 2002-122018/16.
DR   N-PSDB; ABL89669.
XX
```

```
PT   Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT   prevention of neural, immune system, muscular, reproductive,
PT   gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT   disorders -
XX
PS   Claim 11; SEQ ID NO 1636; 2081pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC   (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   147 AA;

  Query Match             8.2%;  Score 591;  DB 23;  Length 147;
  Best Local Similarity   96.6%;  Pred. No. 1.6e-31;
  Matches  115;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         29 MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLA 88

Qy         61 LHLASDFFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLED 119
              ||||||||||| |||||||||||||||||||||||||||||||| ||||||||| || |
Db         89 LHLASDFFLKHPDKDVRLLVACCLADIFRIYAPEAPYTSPDKLXDIFMFITRQLXGLXD 147


RESULT 12
AAY65247
ID   AAY65247 standard; Protein; 101 AA.
XX
AC   AAY65247;
XX
DT   01-FEB-2000  (first entry)
XX
DE   Human 5' EST related polypeptide SEQ ID NO:1408.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW   gene therapy; chromosome mapping; upstream regulatory sequence;
KW   forensic; location; development; protein synthesis; stability;
KW   regulation; identification.
XX
OS   Homo sapiens.
XX
PN   WO9953051-A2.
XX
PD   21-OCT-1999.
XX
PF   09-APR-1999;   99WO-IB00712.
XX
PR   09-APR-1998;   98US-0057719.
PR   28-APR-1998;   98US-0069047.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-038446/03.
DR   N-PSDB; AAZ42861.
```

```
XX
PT   Novel secreted protein 5' expressed sequence tag sequences used in
PT   diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS   Claim 3; Page 776; 837pp; English.
XX
CC   AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
CC   sequences, corresponding to human secreted proteins. AAY64651 to
CC   AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
CC   AAZ43052. The 5' ESTs can be used for producing secreted human gene
CC   products. They can be used to identify and isolate 5' untranslated
CC   regions (UTRs) and upstream regulatory regions which control the
CC   location, development stage, rate, and quantity of protein synthesis, as
CC   well as stability of mRNA. The ESTs are also useful as probes for
CC   chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC   also be used in forensic procedures to identify individuals, or in
CC   diagnostic procedures to identify individuals having genetic diseases
CC   resulting from abnormal gene expression. The products may also be used in
CC   gene therapy protocols. The nucleic acids encoding signal peptides can be
CC   used for directing extracellular secretion of a polypeptide or the
CC   insertion of a polypeptide into a membrane, or importing a polypeptide
CC   into a cell. The proteins encoded by the EST sequences may be useful in
CC   treating a variety of human conditions. Secreted proteins have
CC   therapeutic value, and the identification of new secreted proteins is
CC   valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   101 AA;

  Query Match             7.1%;  Score 513;  DB 21;  Length 101;
  Best Local Similarity   99.0%;  Pred. No. 1.5e-26;
  Matches   99;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLA 60

Qy         61 LHLASDFFLKHPGKDVRLLVACCLADIFRIYAPEAPYTSP 100
              ||||||||||| ||||||||||||||||||||||||||||
Db         61 LHLASDFFLKHPDKDVRLLVACCLADIFRIYAPEAPYTSP 100


RESULT 13
AAG02811
ID   AAG02811 standard; Protein; 92 AA.
XX
AC   AAG02811;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 6892.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J; .
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC02817.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
```

```
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 6892; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   92 AA;

  Query Match             6.5%;  Score 467;  DB 21;  Length 92;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-23;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        839 MKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEI 898
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEI 60

Qy        899 ITLEQYQLCALAINDECYQVRQVFAQKLHKG 929
              |||||||||||||||||||||||||||||||
Db         61 ITLEQYQLCALAINDECYQVRQVFAQKLHKG 91


RESULT 14
AAU27999
ID   AAU27999 standard; Protein; 111 AA.
XX
AC   AAU27999;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human contig polypeptide sequence #152.
XX
KW   Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW   mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW   cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW   nervous system disorder; inflammatory disorder; cell differentiation;
KW   angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW   genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW   cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;
KW   antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW   neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;
KW   immunostimulant; analgesic; gene therapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200164834-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04926.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
PR   17-JUN-2000; 2000US-0597707.
PR   14-JUL-2000; 2000US-0616807.
PR   19-SEP-2000; 2000US-0664641.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Wehrman T,  Wang J,  Ma Y,  Wang D,  Chen R,  Xu C;
PI   Drmanac R;
XX
DR   WPI; 2001-589862/66.
DR   N-PSDB; AAS44899.
XX
PT   Novel polypeptides and nucleic acids obtained from cDNA libraries
PT   prepared from various human tissues, for diagnosis, treatment of
PT   cancer, neurological, inflammatory disorders and for use in arrays for
PT   detection  -
XX
PS   Claim 10; Page 144; 153pp; English.
XX
CC   Sequences AAU27676-AAU28019 represent full-length polypeptides and
CC   contig polypeptides of the invention. The proteins and their associated
CC   DNA sequences are useful for the treatment, diagnosis and prevention of
CC   various types of disorder in a mammalian subject such as a human, dog,
CC   monkey, mouse, hamster or rat. The disorders include cancers such as
CC   leukaemia, lymphoma and neuroblastoma, autoimmune disorders such as
CC   multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC   diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC   disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC   chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC   Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC   disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC   bowel disease. The sequences exhibit activity relating to angiogenesis,
CC   cell proliferation, cell differentiation, stem cell growth factor,
CC   activin or inhibin. Therefore, they can be used to manipulate stem cells
CC   in culture to give rise to neuroepithelial cells that can be used to
CC   augment or replace cells damaged by illness, accidental damage or genetic
CC   disorders. The sequences may also be used for regeneration of bone,
CC   cartilage, tendons and ligaments and in tissue repair and burn healing.
CC   Note: Some sequences for this patent did not form part of the printed
CC   specification, but were obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   111 AA;

  Query Match             3.8%;  Score 276;  DB 22;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-11;
  Matches   55;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1337 TEQKSKSKQHRVSRRAQQRAESPESSAIESTQSTPQKGRGRPSKTPSPSQPKKNV 1391
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TEQKSKSKQHRVSRRAQQRAESPESSAIESTQSTPQKGRGRPSKTPSPSQPKKNV 55


RESULT 15
AAB95449
ID   AAB95449 standard; Protein; 233 AA.
XX
AC   AAB95449;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:17905.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
```

```
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 17905; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   233 AA;

  Query Match             3.4%;  Score 247.5;  DB 22;  Length 233;
  Best Local Similarity   32.3%;  Pred. No. 1.9e-08;
  Matches   81;  Conservative   34;  Mismatches   87;  Indels   49;  Gaps   11;

Qy       1096 RFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETV 1155
              :|||||:|:| | |:|:  | :       ||||  |||||||||: |::    :|: ||
Db          2 KFFTQPEKDFCNDKSYISEETRVLLLTGKPKPAGVLGAVNKPLSATGRKPYVRSTGTETG 61

Qy       1156 S--NASSSSNPSSPGRIKGR-LDSSEMDHSENED--YTMSSPLPGKKSDKRDDSDLVRSE 1210
              |  | :|  |||:  | : :  :::|  ||||:     : |  | |  |
Db         62 SNINVNSELNPSTGNRSREQSSEAAETGVSENEENPVRIISVTPVKNID---------- 110

Qy       1211 LEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESDE-QQWPEEKR 1269
                        || :|  :  |  |      |      | | :||  ||: ::  ||  :||
Db        111 ----------PVKNKE--INSDQAT-----QGNISSDRGKKRTVTAAGAENIQQKTDEK- 152

Qy       1270 LKEDILENEDEQNSP-PKKGKRGRPPKPLG-GGTPKEEPTMKTSKKGSKKKSGPPAPEEE 1327
                      ||   | | | :||| ||     |    | :    |    || |:       |  :|
Db        153 --------VDESGPPAPSKPRRGRRPKSESQGNATKNDDLNKPINKGRKR----AAVGQE 200

Qy       1328 EEEERQSGNTE 1338
                   ::|| :
Db        201 SPGGLEAGNAK 211

Search completed: September 24, 2003, 20:17:11
Job time : 62 secs
```

```
OM nucleic - nucleic search, using sw model

Run on:          September 24, 2003, 09:41:22 ; Search time 8075 Seconds
                                                (without alignments)
                                                12560.072 Million cell updates/sec

Title:           US-09-512-581B-3
Perfect score:   4173
Sequence:        1 atggctcattcaaagactag..........cacaaccaaaaaaaaatgtg 4173

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters:      45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                 1:  em_estba:*
                 2:  em_esthum:*
                 3:  em_estin:*
                 4:  em_estmu:*
                 5:  em_estov:*
                 6:  em_estpl:*
                 7:  em_estro:*
                 8:  em_htc:*
                 9:  gb_est1:*
                 10:  gb_est2:*
                 11:  gb_htc:*
                 12:  gb_est3:*
                 13:  gb_est4:*
                 14:  gb_est5:*
                 15:  em_estfun:*
                 16:  em_estom:*
                 17:  em_gss_hum:*
                 18:  em_gss_inv:*
                 19:  em_gss_pln:*
                 20:  em_gss_vrt:*
                 21:  em_gss_fun:*
                 22:  em_gss_mam:*
                 23:  em_gss_mus:*
                 24:  em_gss_pro:*
                 25:  em_gss_rod:*
                 26:  em_gss_phg:*
                 27:  em_gss_vrl:*
                 28:  gb_gss1:*
                 29:  gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1899.6	45.5	3313	11	AK045159	AK045159 Mus muscu
2	1750.2	41.9	2899	11	AK041682	AK041682 Mus muscu
3	1655.2	39.7	4273	11	AK086753	AK086753 Mus muscu
4	1521	36.4	4427	11	AK032384	AK032384 Mus muscu
5	1095	26.2	1359	11	BC032988	BC032988 Mus muscu
6	716	17.2	791	10	BF509252	BF509252 UI-H-BI4-
7	691.6	16.6	926	14	CA980171	CA980171 AGENCOURT
8	685.8	16.4	745	13	BU940978	BU940978 AGENCOURT
9	677	16.2	692	12	BM794657	BM794657 K-EST0076
10	671	16.1	786	14	CA324097	CA324097 UI-M-FY0-
11	663.4	15.9	734	10	BG256731	BG256731 602370939
12	660.6	15.8	771	14	CB312535	CB312535 AGENCOURT
13	657.2	15.7	695	14	CA405800	CA405800 1001949 H
14	653.8	15.7	786	13	BU703488	BU703488 UI-M-FO0-
15	640.6	15.4	779	14	CA324284	CA324284 UI-M-FY0-
16	634.4	15.2	637	14	CB132602	CB132602 K-EST0183
17	621.2	14.9	848	10	BE873840	BE873840 601483992
18	612.4	14.7	729	10	BG432960	BG432960 602495783
19	608	14.6	624	14	CB215704	CB215704 NISC_np09
20	596	14.3	596	10	BE504550	BE504550 hz58a05.x
21	591.4	14.2	802	10	BG387646	BG387646 602412496
22	588.8	14.1	592	14	CB129904	CB129904 K-EST0179
23	587.8	14.1	782	13	BU363024	BU363024 603790485
24	586.4	14.1	716	14	CB522273	CB522273 UI-M-GH0-
25	577.4	13.8	607	14	CB268446	CB268446 1007352 H
26	562.8	13.5	703	14	CB521463	CB521463 UI-M-GH0-
27	562.6	13.5	908	13	BQ957734	BQ957734 AGENCOURT
28	560.6	13.4	679	10	BB094813	BB094813 BB094813
29	559.2	13.4	578	9	AW770571	AW770571 hl86f11.x
30	548.2	13.1	658	14	CD542155	CD542155 B0238B04-
31	546	13.1	943	13	BU914579	BU914579 AGENCOURT
32	545	13.1	857	14	CA987660	CA987660 AGENCOURT
33	544	13.0	659	12	BI738240	BI738240 603361679
34	535.8	12.8	941	13	BQ886042	BQ886042 AGENCOURT
35	535.4	12.8	931	13	BU914025	BU914025 AGENCOURT
36	532.8	12.8	944	10	BF142489	BF142489 601788573
37	532.4	12.8	882	13	BU916668	BU916668 AGENCOURT
38	531.8	12.7	641	14	CB044718	CB044718 NISC_gc06
39	527.4	12.6	633	14	CD542034	CD542034 B0237B04-
40	526.4	12.6	627	12	BI559686	BI559686 603252554
41	520.2	12.5	532	13	BU429711	BU429711 UI-HF-BN0
42	512.8	12.3	661	10	BB627003	BB627003 BB627003
43	507	12.1	635	10	BB662860	BB662860 BB662860
44	506.2	12.1	686	12	BI851367	BI851367 603377759
45	503.8	12.1	608	9	AV685306	AV685306 AV685306

ALIGNMENTS

```
RESULT 1
AK045159
LOCUS       AK045159               3313 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
            enriched library, clone:B130042B12 product:49J10.1.2
            (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED
            PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo sapiens], full insert
            sequence.
ACCESSION   AK045159
VERSION     AK045159.1  GI:26337114
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
```

```
MEDLINE     20499374
  PUBMED    11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3313)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H.; Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. Tomohiro Kono (Department of Animal
            Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City,
            Kanagawa Prefecture, Japan) whose assistance we gratefully
            acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..3313
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:B130042B12"
                     /db_xref="taxon:10090"
                     /clone="B130042B12"
                     /tissue_type="parthenogenote"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="9.5 days embryo"
     CDS             <1..2476
                     /note="unnamed protein product; 49J10.1.2
                     (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF
                     ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo
                     sapiens] (SPTR|Q96KV4, evidence: FASTY, 92.3%ID,
                     100%length, match=1430)
                     putative"
                     /codon_start=2
                     /protein_id="BAC32242.1"
                     /db_xref="GI:26337115"
                     /translation="VNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAE
                     TFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQA
                     KYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAP
                     LKSLVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNN
                     HSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIIT
                     LEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHA
                     RQCLVKNITVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKD
                     VKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDTKMNEKLYTVCDVAMN
                     IIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTANVLG
                     AVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMS
                     SPLPGKKSDKREDPDLVRSELEKPRSRKKAPVTDPEEKLGMDDLTKLVQEQKPKGSQR
                     GRKRGRTASDSDEQQWPEEKRHKEELLENEDEQNSPPKKGKRGRPPKPLGGGTSKEEP
                     TMKTSKKGNKKKLVPPVVDDDEEEERQIGNTEHKSKSKQHRTSKRAQQRAESPETSAV
                     ESTQSTPQKGRGRPSKAPSPSQPPKKIRVGRSKQVATKENDSSEEMDVLQASSPVSDD
                     TTQEGAEEEDISVGNVRRRSSKRERR"
     polyA_signal    3288..3293
                     /note="putative"
     polyA_site      3313
                     /note="putative"
BASE COUNT     1066 a    709 c    732 g    806 t
ORIGIN

  Query Match             45.5%;  Score 1899.6;  DB 11;  Length 3313;
  Best Local Similarity   89.2%;  Pred. No. 0;
  Matches 2049;  Conservative    0;  Mismatches  249;  Indels    0;  Gaps    0;

Qy       1872 AGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTTCCAACTGATCA 1931
              |||||| ||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db          1 AGTGAATAAATCAATAGATGGAACAGCAGATGATGAGGATGAGGGTGTTCCAACTGATCA 60

Qy       1932 AGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACATCCCATCTCATT 1991
              ||| ||||| ||||| ||||| ||||| ||||| || || ||||| || || ||||| ||
Db         61 AGCTATCAGGGCAGGCCTTGAGCTGCTGAAGGTGCTGTCTTTTACGCACCCTATCTCCTT 120

Qy       1992 TCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGATGATGAAAAAGT 2051
              ||| |||||||| || ||||| ||  | |||||||||||||| ||||||||||| || ||
Db        121 TCACTCTGCTGAGACGTTTGAGTCCCTTCTGGCTTGTCTGAAGATGGATGATGAGAAGGT 180

Qy       2052 AGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAAGAGGATTTTCC 2111
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        181 GGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAAGAGGATTTCCC 240

Qy       2112 ACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAAGGACCCCCCCG 2171
               ||||||||||||||||||||||||||  ||||||||||||| ||||||||||| || ||
Db        241 TCACATCAGATCAGCCTTGCTTCCTGTACTACATCACAAATCCAAAAAAGGACCACCTCG 300

Qy       2172 TCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAAGAGACCCAGTT 2231
```

```
                 || |||||||| ||||| |||||||| ||||| |||||||||||||||| ||||||||
Db        301 GCAGGCCAAATACGCCATCCATTGTATTCATGCCATATTTCTAGTAAAGAAACCCAGTT 360

Qy       2232 TGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTGGAACATCTCAT 2291
              ||||||||||||||||||||||||| |||||||||||||||||||||| ||||| ||
Db        361 TGCACAGATATTTGAGCCTCTGCATAAAAGCCTAGATCCAAGCAACCTGGAGCATCTTAT 420

Qy       2292 AACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAATTTGCTGCTCC 2351
              ||| ||  |||| ||||||||||||||||||| || ||||||||||| || ||||||||
Db        421 AACCCCTCTGGTCACTATTGGTCATATTGCTCTTCTCGCACCTGATCAGTTCGCTGCTCC 480

Qy       2352 TTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAATGATCGGCTTCC 2411
              |  ||| |||| ||| || ||||||||||||| || |||||||||| ||||||||
Db        481 TCTGAAGTCTTTGGTGGCAACTTTCATTGTGAAGGACCTCCTCATGAATGACCGGCTTCC 540

Qy       2412 AGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCTGAGACAATGGT 2471
              ||| |||||||||||||| ||||||||||||||||| ||||| || ||||||||||||||
Db        541 AGGAAAAAAGACAACTAAGCTTTGGGTTCCAGATGAGGAAGTCTCACCTGAGACAATGGT 600

Qy       2472 CAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAAAATAATCACAG 2531
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        601 CAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAGAATAATCACAG 660

Qy       2532 TAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGTGATGGAGACTT 2591
              ||| ||||||||||| ||| | ||||| |||||||| ||| ||||||||||||| || ||
Db        661 TAAGTCAGGAACTTCCACCCTCAGATTACTAACAACGATACTGCATAGTGATGGGGATTT 720

Qy       2592 GACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTTGCTGCTGGGAG 2651
              |||||||| |||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        721 GACAGAACAAGGGAAAATTAGTAAACCAGATATGTCACGCCTGAGACTTGCTGCTGGGAG 780

Qy       2652 TGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACATTAGAACAATA 2711
              |||||||||||||||||||| || ||||||||||| || ||||| ||| | || || ||
Db        781 TGCTATTGTGAAGCTGGCACAGGAGCCCTGTTACCACGAGATCATTACACTGGAGCAGTA 840

Qy       2712 TCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAAGTGTTTGCCCA 2771
               ||||| ||||||||||| ||||| ||||| || |||||||| || || ||||| || ||
Db        841 CCAGCTGTGTGCATTAGCCATCAATGATGAGTGTTATCAAGTCAGGCAGGTGTTCGCTCA 900

Qy       2772 GAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATGGCAATCTGTGC 2831
              |||||||||||||||||||||| |||||||||||| |||||||| ||||| ||||||||
Db        901 GAAACTTCACAAAGGCCTTTCCCGCTTACGGCTTCCCCTTGAGTACATGGCCATCTGTGC 960

Qy       2832 CCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAATGTTTGGTGAA 2891
               ||||||| ||||| |||||||| || |||||    |||||||| || ||||||||
Db        961 TCTTTGTGCCAAAGACCCTGTGAAAGAGAGGCGAGCCCATGCTAGACAGTGCTTGGTGAA 1020

Qy       2892 AAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGTGAAAAATTATT 2951
               || || | ||| ||| ||||||| ||||||||||||||||||||||||||||||||||
Db       1021 GAACATCACTGTGAGGAGGGAGTACCTGAAGCAGCATGCAGCTGTTAGTGAAAAATTATT 1080

Qy       2952 GTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCACATGACCCAGA 3011
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCACATGACCCAGA 1140

Qy       3012 TTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGTCTTTGGTTTGT 3071
              |||||||||||||||||||||||||||||||||||||| ||||||||| ||||||||||
Db       1141 TTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTGAAAGAATGCCTTTGGTTTGT 1200

Qy       3072 TCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATCAGAAAGATGGT 3131
              |||||| ||||| |||||||||||||||| || ||  |||||||||||| ||||||
Db       1201 TCTGGAGATATTGATGGCTAAAAATGAAAACAACAGCCATGCATTTATCAGAAAAATGGT 1260

Qy       3132 AGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAAATGAATGAAAA 3191
              |||||||||||||| ||||||||||| |||||||||||||||| |||||||||||||||
Db       1261 AGAAAATATTAAACAGACAAAAGATGCTCAAGGACCAGATGATACAAAAATGAATGAAAA 1320

Qy       3192 ACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGTACTACATACAG 3251
              | ||||||| |||||||||||||||||||| ||||||||||| ||||| || || |||||
Db       1321 ATTGTACACCGTGTGTGATGTTGCCATGAACATCATCATGTCCAAGAGCACCACGTACAG 1380

Qy       3252 TTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAACCTGACAAGAA 3311
               |||| |||||||| ||||| || || |||||||| || ||||| || |||||||||||
Db       1381 CCTGGAGTCTCCTAAGGACCCCGTGCTGCCAGCTCGGTTTTTCACCCAGCCTGACAAGAA 1440

Qy       3312 TTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTCACTCCTGGAAA 3371
              ||| ||||||||||||||||| |||||||| |||||||||||||||||||||||||||||
Db       1441 TTTTAGTAACACCAAAAATTACCTGCCTCCAGAAATGAAATCATTTTTCACTCCTGGAAA 1500

Qy       3372 ACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCAGCAGGCAAGCA 3431
              ||||||||| ||||||||| ||||| ||||| |||||||||||||||||||||| ||
Db       1501 ACCTAAAACAGCCAATGTTCTCGGAGCAGTTAATAAGCCACTTTCATCAGCAGGCAAACA 1560

Qy       3432 ATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGCAGCTCAAATCC 3491
               |||||||||||||||| |||||||||||| ||||| |||||||||||||| || ||
Db       1561 GTCTCAGACCAAATCATCAAGAATGGAAACTGTGAGCAACGCAAGCAGCAGCTCCAACCC 1620

Qy       3492 AAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGATCACAGTGAAAA 3551
              |||||||||||||| || ||||||||||||||||||| ||||||||||||||||||||||
Db       1621 AAGCTCTCCTGGAAGGATCAAGGGGAGGCTTGATAGCTCTGAAATGGATCACAGTGAAAA 1680

Qy       3552 TGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAGAGAGACGACTC 3611
              |||||||| ||||||||||||||||||| || ||||||||||||||||||||| ||| |
Db       1681 TGAAGATTATACAATGTCTTCACCTTTGCCAGGAAAAAAAAGTGACAAGAGAGAAGACCC 1740

Qy       3612 TGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACGCCCGTCACAGA 3671
              |||||||||||||||| |||||||||||||| |  |||||||| | || |||||||
Db       1741 TGATCTTGTAAGGTCTGAGTTGGAGAAGCCTAGAAGTCGGAAAAAAGCACCTGTCACAGA 1800

Qy       3672 ACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAACAGAAACCTAA 3731
                |  || |||||||||||||||||||| | ||||||||||||||||||||||||||||
Db       1801 CCCTGAAGAGAAATTAGGTATGGATGACCTAACTAAGTTGGTACAGGAACAGAAACCTAA 1860

Qy       3732 AGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGTG 3791
              |||||||||||| | |||||||||||| || ||||||| | || || || ||||||||
Db       1861 AGGCAGTCAGCGAGGCCGGAAAAGAGGCCGTACAGCTTCAGACTCAGACGAGCAGCAGTG 1920

Qy       3792 GCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCC 3851
              |||||||| ||||||| |||||| ||   |  |||||||| |||| ||||| |||| ||
Db       1921 GCCTGAGGAGAAGAGGCACAAAGAGGAGCTCCTGGAAAATGAGGATGAGCAGAACAGCCC 1980

Qy       3852 GCCAAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGA 3911
               |||||||||| |||||||| | ||||||||||||||||||| || ||| | || ||
Db       1981 ACCAAAAAAAGGGCAAAAGAGGCAGGCCACCAAAACCTCTTGGTGGGGGGACATCGAAGGA 2040

Qy       3912 AGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACC 3971
              |||||| |||||||||||| || || ||||||| ||| ||||||  ||||||| | |
Db       2041 AGAGCCCACAATGAAAACATCCAAGAAAGGAAACAAGAAAAAACTTGTACCTCCTGTAGT 2100

Qy       3972 AGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAA 4031
              ||| || || ||||||||||||||||||| |||| ||||| |||| ||||||||||||
Db       2101 AGACGACGATGAAGAAGAAGAAAGACAAATTGGAAACACAGAACATAAGTCAAAAAGCAA 2160

Qy       4032 ACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAATCTAGTGCAAT 4091
              ||||||||       |||| ||||||||| |||||||||||||||||||| | |||||| |
Db       2161 ACAGCACCGCACATCAAAGAGAGCACAACAGAGAGCAGAATCTCCTGAAACAAGTGCAGT 2220

Qy       4092 TGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAAACGCCATCACC 4151
              |||||||||||||||||||||||||||||||||||||||||||||||||| | ||||||||
Db       2221 TGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAAGCACCATCACC 2280

Qy       4152 ATCACAACCAAAAAAAAA 4169
              |||||||||    |||||
Db       2281 ATCACAACCCCCCAAAAA 2298
```

```
RESULT 2
AK041682
LOCUS       AK041682               2899 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
            library, clone:A630029M15 product:49J10.1.2 (ANDROGEN-INDUCED
            PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2)
            (FRAGMENT) homolog [Homo sapiens], full insert sequence.
ACCESSION   AK041682
VERSION     AK041682.1  GI:26334660
KEYWORDS    HTC; CAP trapper.
```

```
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2899)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
            Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
            Trust/MRC building Addenbrookes Hospital Cambridge) whose
            assistance we gratefully acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..2899
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:A630029M15"
                     /db_xref="taxon:10090"
                     /clone="A630029M15"
                     /tissue_type="thymus"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="3 days neonate"
     CDS             117..2222
                     /note="unnamed protein product; 49J10.1.2
                     (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF
                     ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo
                     sapiens] (SPTR|Q96KV4, evidence: FASTY, 92.3%ID,
                     100%length, match=1430)
                     putative"
                     /codon_start=1
                     /protein_id="BAC31031.1"
                     /db_xref="GI:26334661"
                     /translation="MVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEALVSPTCSC
                     KQAEGCVREITKKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQVNKSIDG
                     TADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAAL
                     QIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQI
                     FEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPLKSLVATFIVKDLLMNDRLPG
                     KKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGD
                     LTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQV
                     FAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNITVRREYLKQHAAV
                     SEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSH
                     AFIRKMVENIKQTKDAQGPDDTKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPA
                     RFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTANVLGAVNKPLSSAGKQSQTKSSRME
                     TVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKREDPDLVRVRC
                     LVGRVMRLLIVIVLVIFAL"
BASE COUNT      868 a    582 c    617 g    832 t
ORIGIN

  Query Match            41.9%;  Score 1750.2;  DB 11;  Length 2899;
  Best Local Similarity  90.2%;  Pred. No. 4.8e-290;
  Matches 1872;  Conservative    0;  Mismatches  203;  Indels    0;  Gaps    0;

Qy       1550 AAACAGATGCCAGTGTCAAGGCCATATTTCAAAAGTGATGGTTATTACAAGAAATTTAC 1609
              | |||||||||||||||||||||||||||| |||||||||||||| ||||||||||| |
Db         79 AGACAGATGCCAGTGTCAAGGCCATATTTCTAAAGTGATGGTTATCACAAGAAATTTGC 138

Qy       1610 CTGATCCTGGTAAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATG 1669
              | |||||||||||||| || || ||||||||||| |||||||| ||||| ||||||||||
Db        139 CAGATCCTGGTAAGGCCCAAGACTTCATGAAGAAGTTCACACAAGTGTTGGAAGATGATG 198

Qy       1670 AGAAAATAAGAAAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTG 1729
              |||||||| | || ||| | ||||  ||||||||||| ||||| ||||||||||||||||
Db        199 AGAAAATACGGAAACAGCTGGAAGCCCTTGTTAGTCCCACATGTTCCTGCAAGCAGGCTG 258
```

```
Qy   1730 AAGGTTGTGTGCGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTT 1789
Db    259 AAGGCTGTGTGCGTGAAATAACTAAGAAGTTGGGCAACCCTAAGCAGCCTACAAACCCTT 318

Qy   1790 TCCTGGAAATGATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAAT 1849
Db    319 TCTTGGAAATGATCAAGTTTCTCTTGGAGAGGATAGCTCCTGTACACATAGACACTGAAT 378

Qy   1850 CTATCAGTGCTCTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAG 1909
Db    379 CCATCAGTGCTCTTATCAAGCAAGTGAATAAATCAATAGATGGAACAGCAGATGATGAGG 438

Qy   1910 ATGAGGGTGTTCCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCT 1969
Db    439 ATGAGGGTGTTCCAACTGATCAAGCTATCAGGGCAGGCCTTGAGCTGCTGAAGGTGCTGT 498

Qy   1970 CATTTACACATCCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTC 2029
Db    499 CTTTTACGCACCCTATCTCCTTTCACTCTGCTGAGACGTTTGAGTCCCTTCTGGCTTGTC 558

Qy   2030 TGAAAATGGATGATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAA 2089
Db    559 TGAAGATGGATGATGAGAAGGTGGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAA 618

Qy   2090 GCAAAATTGAAGAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACA 2149
Db    619 GCAAAATTGAAGAGGATTTCCCTCACATCAGATCAGCCTTGCTTCCTGTACTACATCACA 678

Qy   2150 AATCTAAAAAAGGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATAT 2209
Db    679 AATCCAAAAAAGGACCACCTCGGCAGGCCAAATACGCCATCCATTGTATTCATGCCATAT 738

Qy   2210 TTTCTAGTAAAGAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATC 2269
Db    739 TTTCTAGTAAAGAAACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAAAGCCTAGATC 798

Qy   2270 CAAGCAACCTGGAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTG 2329
Db    799 CAAGCAACCTGGAGCATCTTATAACCCCTCTGGTCACTATTGGTCATATTGCTCTTCTCG 858

Qy   2330 CACCTGATCAATTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATC 2389
Db    859 CACCTGATCAGTTCGCTGCTCCTCTGAAGTCTTTGGTGGCAACTTTCATTGTGAAGGACC 918

Qy   2390 TTCTCATGAATGATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAG 2449
Db    919 TCCTCATGAATGACCGGCTTCCAGGAAAAAAGACAACTAAGCTTTGGGTTCCAGATGAGG 978

Qy   2450 AAGTATCTCCTGAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTAC 2509
Db    979 AAGTCTCACCTGAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTAC 1038

Qy   2510 TTGGAATGAAAAATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAA 2569
Db   1039 TTGGAATGAAGAATAATCACAGTAAGTCAGGAACTTCCACCCTCAGATTACTAACAACGA 1098

Qy   2570 TATTGCATAGTGATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCAC 2629
Db   1099 TACTGCATAGTGATGGGGATTTGACAGAACAAGGGAAAATTAGTAAACCAGATATGTCAC 1158

Qy   2630 GTCTGAGACTTGCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATG 2689
Db   1159 GCCTGAGACTTGCTGCTGGGAGTGCTATTGTGAAGCTGGCACAGGAGCCCTGTTACCACG 1218

Qy   2690 AAATCATCACATTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATC 2749
Db   1219 AGATCATTACACTGGAGCAGTACCAGCTGTGTGCATTAGCCATCAATGATGAGTGTTATC 1278

Qy   2750 AAGTAAGACAAGTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCAC 2809
Db   1279 AAGTCAGGCAGGTGTTCGCTCAGAAACTTCACAAAGGCCTTTCCCGCTTACGGCTTCCCC 1338

Qy   2810 TTGAGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTC 2869
Db   1339 TTGAGTACATGGCCATCTGTGCTCTTTGTGCCAAAGACCCTGTGAAAGAGAGGCGAGCCC 1398

Qy   2870 ATGCTAGGCAATGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATG 2929
Db   1399 ATGCTAGACAGTGCTTGGTGAAGAACATCACTGTGAGGAGGGAGTACCTGAAGCAGCATG 1458

Qy   2930 CAGCTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTC 2989
Db   1459 CAGCTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTC 1518

Qy   2990 ACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATG 3049
Db   1519 ACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATG 1578

Qy   3050 TTAAAGAATGTCTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTC 3109
Db   1579 TGAAAGAATGCCTTTGGTTTGTTCTGGAGATATTGATGGCTAAAAATGAAAACAACAGCC 1638

Qy   3110 ACGCTTTTATCAGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAG 3169
Db   1639 ATGCATTTATCAGAAAAATGGTAGAAAATATTAAACAGACAAAAGATGCTCAAGGACCAG 1698

Qy   3170 ATGATGCAAAAATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCA 3229
Db   1699 ATGATACAAAAATGAATGAAAAATTGTACACCGTGTGTGATGTTGCCATGAACATCATCA 1758

Qy   3230 TGTCAAAGAGTACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTT 3289
Db   1759 TGTCCAAGAGCACCACGTACAGCCTGGAGTCTCCTAAGGACCCCGTGCTGCCAGCTCGGT 1818

Qy   3290 TCTTCACTCAACCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGA 3349
Db   1819 TTTTCACCCAGCCTGACAAGAATTTTAGTAACACCAAAAATTACCTGCCTCCAGAAATGA 1878

Qy   3350 AATCATTTTTCACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGC 3409
Db   1879 AATCATTTTTCACTCCTGGAAAACCTAAAACAGCCAATGTTCTCGGAGCAGTTAATAAGC 1938

Qy   3410 CACTTTCATCAGCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCA 3469
Db   1939 CACTTTCATCAGCAGGCAAACAGTCTCAGACCAAATCATCAAGAATGGAAACTGTGAGCA 1998

Qy   3470 ATGCAAGCAGCAGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTT 3529
Db   1999 ACGCAAGCAGCAGCTCCAACCCAAGCTCTCCTGGAAGGATCAAGGGGAGGCTTGATAGCT 2058

Qy   3530 CTGAAATGGATCACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAA 3589
Db   2059 CTGAAATGGATCACAGTGAAAATGAAGATTATACAATGTCTTCACCTTTGCCAGGAAAAA 2118

Qy   3590 AAAGTGACAAGAGAGACGACTCTGATCTTGTAAGG 3624
Db   2119 AAAGTGACAAGAGAGAAGACCCTGATCTTGTAAGG 2153
```

RESULT 3
AK086753

```
LOCUS       AK086753                4273 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 15 days embryo head cDNA, RIKEN full-length enriched
            library, clone:D930049H23 product:49J10.1.2 (ANDROGEN-INDUCED
            PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2)
            (FRAGMENT) homolog [Homo sapiens], full insert sequence.
ACCESSION   AK086753
VERSION     AK086753.1  GI:26103739
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
```

```
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 4273)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
```

```
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..4273
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:D930049H23"
                     /db_xref="taxon:10090"
                     /clone="D930049H23"
                     /tissue_type="head"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="15 days embryo"
     misc_feature    <1..2245
                     /note="49J10.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE
                     SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog
                     [Homo sapiens] (SPTR|Q96KV4, evidence: FASTY, 92.3%ID,
                     100%length, match=1430)
                     putative"
BASE COUNT     1363 a    840 c    899 g   1171 t
ORIGIN

  Query Match             39.7%;  Score 1655.2;  DB 11;  Length 4273;
  Best Local Similarity   88.8%;  Pred. No. 9.2e-274;
  Matches 1836;  Conservative    0;  Mismatches  228;  Indels    4;  Gaps    4;

Qy       2104 GATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCT-AAAAAAGG 2162
              ||||| || ||||||||||||||||||||||||||  ||||||||||||||  |||||||
Db          1 GATTTCCCTCACATCAGATCAGCCTTGCTTCCTGTACTACATCACAAATCCAAAAAAAGG 60

Qy       2163 ACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAAGA 2222
              ||| || ||  | | ||||||| ||||| |||||||| ||||| |||||||||||||||||
Db         61 ACCACCTCGGCAGGCCAAATACGCCATCCATTGTATTCATGCCATATTTTCTAGTAAAGA 120

Qy       2223 GACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTGGA 2282
               ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db        121 AACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAAAGCCTAGATCCAAGCAACCTGGA 180

Qy       2283 ACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAATT 2342
               ||||| ||||| ||  |||| |||||||||||||||||||| || ||||||||||| ||
Db        181 GCATCTTATAACCCCTCTGGTCACTATTGGTCATATTGCTCTTCTCGCACCTGATCAGTT 240

Qy       2343 TGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAATGA 2402
               |||||||||  ||| |||| ||| || |||||||||||||| || || |||||||||||
Db        241 CGCTGCTCCTCTGAAGTCTTTGGTGGCAACTTTCATTGTGAAGGACCTCCTCATGAATGA 300

Qy       2403 TCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCTGA 2462
               ||||||||||| |||||||||||||| ||||||||||||||||| ||||| || |||||
Db        301 CCGGCTTCCAGGAAAAAAGACAACTAAGCTTTGGGTTCCAGATGAGGAAGTCTCACCTGA 360

Qy       2463 GACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAAAA 2522
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        361 GACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAGAA 420

Qy       2523 TAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGTGA 2582
              |||||||||||| ||||||||||| ||| | ||||| |||||||| ||| ||||||||||
Db        421 TAATCACAGTAAGTCAGGAACTTCCACCCTCAGATTACTAACAACGATACTGCATAGTGA 480

Qy       2583 TGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTTGC 2642
              ||| || ||||||||||| ||||||||||||||||||||||||||||| |||||||||||
Db        481 TGGGGATTTGACAGAACAAGGGAAAATTAGTAAACCAGATATGTCACGCCTGAGACTTGC 540

Qy       2643 TGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACATT 2702
              |||||||||||||||||||||||||||||| || ||||||||||| || ||||| ||| |
```

```
Db        541 TGCTGGGAGTGCTATTGTGAAGCTGGCACAGGAGCCCTGTTACCACGAGATCATTACACT 600

Qy       2703 AGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAAGT 2762
Db        601 GGAGCAGTACCAGCTGTGTGCATTAGCCATCAATGATGAGTGTTATCAAGTCAGGCAGGT 660

Qy       2763 GTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTT-GAGTATATGG 2821
Db        661 GTTCGCTCAGAAACTTCACAAAGGCCTTTCCCGCTTACGGCTTCCCCTTGGAGTACATGG 720

Qy       2822 CAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAAT 2881
Db        721 CCATCTGTGCTCTTTGTGCCAAAGACCCTGTGAAAGAGAGGGGAGCCCATGCTAGACAGT 780

Qy       2882 GTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGTG 2941
Db        781 GCTTGGTGAAGAACAT-CCCGTGAGGAGGGAGTACCTGAAGCAGCATGCAGCTGTTAGTG 839

Qy       2942 AAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCAC 3001
Db        840 AAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCAC 899

Qy       3002 ATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGTC 3061
Db        900 ATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTGAAAGAATGCC 959

Qy       3062 TTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATCA 3121
Db        960 TTTGGTTTGTTCTGGAGATATTGATGGCTAAAAATGAAAACAACAGCCATGCATTTATCA 1019

Qy       3122 GAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAAA 3181
Db       1020 GAAAAATGGTAGAAAATATTAAACAGACAAAAGATGCTCAAGGACCAGATGATACAAAAA 1079

Qy       3182 TGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGTA 3241
Db       1080 TGAATGAAAAATTGTACACCGTGTGTGATGTTGCCATGAACATCATCATGTCCAAGAGCA 1139

Qy       3242 CTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAAC 3301
Db       1140 CCACGTACAGCCTGGAGTCTCCTAAGGACCCCGTGCTGCCAGCTCGGTTTTCACCCAGC 1199

Qy       3302 CTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTCA 3361
Db       1200 CTGACAAGAATTTTAGTAACACCAAAAATTACCTGCCTCCAGAAATGAAATCATTTTTCA 1259

Qy       3362 CTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCAG 3421
Db       1260 CTCCTGGAAAACCTAAAACAGCCAATGTTCTCGGAGCAGTTAATAAGCCACTTTCATCAG 1319

Qy       3422 CAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGCA 3481
Db       1320 CAGGCAAACAGTCTCAGACCAAATCATCAAGAATGGAAACTGTGAGCAACGCAAGCAGCA 1379

Qy       3482 GCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGATC 3541
Db       1380 GCTCCAACCCAAGCTCTCCTGGAAGGAT-CAGGGGAGGCTTGATAGCTCTGAAATGGATC 1438

Qy       3542 ACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAGA 3601
Db       1439 ACAGTGAAAATGAAGATTATACAATGTCTTCACCTTTGCCAGGAAAAAAAAGTGACAAGA 1498

Qy       3602 GAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACGC 3661
Db       1499 GAGAAGACCCTGATCTTGTAAGGTCTGAGTTGGAGAAGCCTAGAAGTCGGAAAAAAGCAC 1558

Qy       3662 CCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAAC 3721
Db       1559 CTGTCACAGACCCTGAAGAGAAATTAGGTATGGATGACCTAACTAAGTTGGTACAGGAAC 1618

Qy       3722 AGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGATG 3781
Db       1619 AGAAACCTAAAGGCAGTCAGCGAGGCCGGAAAAGAGGCCGTACAGCTTCAGACTCAGACG 1678

Qy       3782 AACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAAC 3841
Db       1679 AGCAGCAGTGGCCTGAGGAGAAGAGGCACAAAGAGGAGCTCCTGGAAAATGAGGATGAGC 1738

Qy       3842 AGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTA 3901
Db       1739 AGAACAGCCCACCAAAAAAGGGCAAAAGAGGCAGGCCACCCAAACCTCTTGGTGGGGGA 1798

Qy       3902 CACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGAC 3961
Db       1799 CATCGAAGGAAGAGCCCACAATGAAAACATCCAAGAAAGGAAACAAGAAAAAACTTGTAC 1858

Qy       3962 CTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGT 4021
Db       1859 CTCCTGTAGTAGACGACGATGAAGAACAAGAAAGACAAATTGGAAACACAGAACATAAGT 1918

Qy       4022 CCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAAT 4081
Db       1919 CAAAAAGCAAACAGCACCGCACATCAAAGAGAGCACAACAGAGAGCAGAATCTCCTGAAA 1978

Qy       4082 CTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAAA 4141
Db       1979 CAAGTGCAGTTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAAG 2038

Qy       4142 CGCCATCACCATCACAACCAAAAAAAAA 4169
Db       2039 CACCATCACCATCACAACCCCCCCAAAA 2066
```

RESULT 4
AK032384
LOCUS AK032384 4427 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length enriched library, clone:6430530P15 product:49J10.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog [Homo sapiens], full insert sequence.
ACCESSION AK032384
VERSION AK032384.1 GI:26082814
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4

```
   AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
             Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
             Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
             Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
             Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
             Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
             Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
             Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
             Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
             Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
             Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
             Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
             Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
             Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
             Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
             Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
             Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
             Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
             and Hayashizaki,Y.
   TITLE     Functional annotation of a full-length mouse cDNA collection
   JOURNAL   Nature 409 (6821), 685-690 (2001)
   MEDLINE   21085660
    PUBMED   11217851
REFERENCE    5
   AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
             Group Phase I & II Team.
   TITLE     Analysis of the mouse transcriptome based on functional annotation
             of 60,770 full-length cDNAs
   JOURNAL   Nature 420, 563-573 (2002)
REFERENCE    6  (bases 1 to 4427)
   AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
             Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
             Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
             Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
             Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
             Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
             Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
             Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
             Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
             Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
             Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
             Muramatsu,M. and Hayashizaki,Y.
   TITLE     Direct Submission
   JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
             Physical and Chemical Research (RIKEN), Laboratory for Genome
             Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
             RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
             Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
             URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
             Fax:81-45-503-9216)
COMMENT      cDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in RIKEN.
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissues.
             Please visit our web site for further details.
             URL:http://genome.gsc.riken.go.jp/
             URL:http://fantom.gsc.riken.go.jp/.
FEATURES              Location/Qualifiers
     source           1. .4427
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6J"
                      /db_xref="FANTOM_DB:6430530P15"
                      /db_xref="taxon:10090"
                      /clone="6430530P15"
                      /sex="male"
                      /tissue_type="olfactory brain"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     misc_feature     <1. .2062
                      /note="49J10.1.2 (ANDROGEN-INDUCED PROSTATE PROLIFERATIVE
                      SHUTOFF ASSOCIATED PROTEIN, ISOFORM 2) (FRAGMENT) homolog
```

```
                      [Homo sapiens] (SPTR|Q96KV4, evidence: FASTY, 92.3%ID,
                      100%length, match=1430)
                      putative"
     polyA_signal     4406. .4411
                      /note="putative"
     polyA_site       4427
                      /note="putative"
BASE COUNT     1423 a    832 c    919 g   1253 t
ORIGIN

  Query Match             36.4%;  Score 1521;  DB 11;  Length 4427;
  Best Local Similarity   88.5%;  Pred. No. 9.4e-251;
  Matches 1674;  Conservative    0;  Mismatches  210;  Indels    7;  Gaps    2;

Qy       2279 TGGAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATC 2338
              |||| ||||| ||||| ||  |||| |||||||||||||||||||||| || ||||||||||||
Db          1 TGGAGCATCTTATAACCCCTCTGGTCACTATTGGTCATATTGCTCTTCTCGCACCTGATC 60

Qy       2339 AATTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGA 2398
              | || |||||||||  ||| |||| ||| || |||||||||||||||| || || |||||||
Db         61 AGTTCGCTGCTCCTCTGAAGTCTTTGGTGGCAACTTTCATTGTGAAGGACCTCCTCATGA 120

Qy       2399 ATGATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTC 2458
              |||| ||||||||||| ||||||||||||||| ||||||||||||||||| ||||| || |
Db        121 ATGACCGGCTTCCAGGAAAAAAGACAACTAAGCTTTGGGTTCCAGATGAGGAAGTCTCAC 180

Qy       2459 CTGAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGA 2518
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTGAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGA 240

Qy       2519 AAAATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATA 2578
              | |||||||||||||| ||||||||||| ||| | ||||| |||||||| ||| ||||||
Db        241 AGAATAATCACAGTAAGTCAGGAACTTCCACCCTCAGATTACTAACAACGATACTGCATA 300

Qy       2579 GTGATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGAC 2638
              ||||||| || |||||||||||| ||||||||||||||||||||||||||||| |||||||
Db        301 GTGATGGGGATTTGACAGAACAAGGGAAAATTAGTAAACCAGATATGTCACGCCTGAGAC 360

Qy       2639 TTGCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCA 2698
              |||||||||||||||||||||||||||||||||| || ||||||||||| || ||||| |
Db        361 TTGCTGCTGGGAGTGCTATTGTGAAGCTGGCACAGGAGCCCTGTTACCACGAGATCATTA 420

Qy       2699 CATTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGAC 2758
              || | || || || ||||| ||||||||||| ||||| ||||| || |||||||| || |
Db        421 CACTGGAGCAGTACCAGCTGTGTGCATTAGCCATCAATGATGAGTGTTATCAAGTCAGGC 480

Qy       2759 AAGTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATA 2818
              | ||||| || ||||||||||||||||||||||||| ||||||||||| |||||||| |
Db        481 AGGTGTTCGCTCAGAAACTTCACAAAGGCCTTTCCCGCTTACGGCTTCCCCTTGAGTACA 540

Qy       2819 TGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGC 2878
              |||| |||||||| |||||||| ||||| ||||| || |||||  |||| |||||||| |
Db        541 TGGCCATCTGTGCTCTTTGTGCCAAAGACCCTGTGAAAGAGAGGCGAGCCCATGCTAGAC 600

Qy       2879 AATGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTA 2938
              | || |||||||| || || | ||| ||| ||||||| ||||||||||||||||||||||
Db        601 AGTGCTTGGTGAAGAACATCACTGTGAGGAGGGAGTACCTGAAGCAGCATGCAGCTGTTA 660

Qy       2939 GTGAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGG 2998
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTGAAAAATTAATGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGG 720

Qy       2999 CACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAAT 3058
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        721 CACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTGAAAGAAT 780

Qy       3059 GTCTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTA 3118
              | ||||||||||||||||| ||||| ||||||||||||||||| ||||| || || ||||
Db        781 GCCTTTGGTTTGTTCTGGAGATATTGATGGCTAAAAATGAAAACAACAGCCATGCATTTA 840

Qy       3119 TCAGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAA 3178
              ||||||| |||||||||||||||||||| ||||||||||| |||||||||||||||| |||
Db        841 TCAGAAAAATGGTAGAAAATATTAAACAGACAAAAGATGCTCAAGGACCAGATGATACAA 900
```

```
Qy       3179 AAATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGA 3238
              ||||||||||||| |||||| ||||||||||||||||||||||| ||||||||||| ||||
Db        901 AAATGAATGAAAAATTGTACACCGTGTGTGATGTTGCCATGAACATCATCATGTCCAAGA 960

Qy       3239 GTACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTC 3298
              | || || |||||  |||| |||||||| ||||| || || |||||||| || ||||| |
Db        961 GCACCACGTACAGCCTGGAGTCTCCTAAGGACCCCGTGCTGCCAGCTCGGTTTTTCACCC 1020

Qy       3299 AACCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTT 3358
               ||||||||||||| |||||||||||||||| |||||||| ||||||||||||||||||
Db       1021 AGCCTGACAAGAATTTTAGTAACACCAAAAATTACCTGCCTCCAGAAATGAAATCATTTT 1080

Qy       3359 TCACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCAT 3418
              |||||||||||||||||||||| |||||||||| ||||| | ||| ||||||||||||
Db       1081 TCACTCCTGGAAAACCTAAAACAGCCAATGTTCTCGGAGCAG-TAATAAGCCACTTTCAT 1139

Qy       3419 CAGCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCA 3478
              ||||||||| || ||||||||||||||||| |||||||||||| ||||| |||||||
Db       1140 CAGCAGGCAAACAGTCTCAGACCAAATCATCAAGAATGGAAACTGTGAGCAACGCAAGCA 1199

Qy       3479 GCAGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGG 3538
              |||||| || |||||||||||||||| || |||||||||||||||| ||||||||||
Db       1200 GCAGCTCCAACCCAAGCTCTCCTGGAAGGATCAAGGGGAGGCTTGATAGCTCTGAAATGG 1259

Qy       3539 ATCACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACA 3598
              |||||||||||||||||||| |||||||||||||||||||| || ||||||||||||
Db       1260 ATCACAGTGAAAATGAAGATTATACAATGTCTTCACCTTTGCCAGGAAAAAAAAGTGACA 1319

Qy       3599 AGAGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAA 3658
              |||||| ||| ||||||||        ||||| ||||||||||||||| |   ||||||||
Db       1320 AGAGAGAAGACCCTGATCTT------TCTGAGTTGGAGAAGCCTAGAAGTCGGAAAAAAG 1373

Qy       3659 CGCCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGG 3718
               | || |||||||| |    || ||||||||||||||||||||| | |||||||||||||
Db       1374 CACCTGTCACAGACCCTGAAGAGAAATTAGGTATGGATGACCTAACTAAGTTGGTACAGG 1433

Qy       3719 AACAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTG 3778
              ||||||||||||||||||||||||||| | ||||||||||||| ||| |||||||| || |
Db       1434 AACAGAAACCTAAAGGCAGTCAGCGAGGCCGGAAAAGAGGCCGTACAGCTTCAGACTCAG 1493

Qy       3779 ATGAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATG 3838
              | || |||||||||||||||| ||||||| |||||| ||   |  | ||||||| ||||
Db       1494 ACGAGCAGCAGTGGCCTGAGGAGAAGAGGCACAAAGAGGAGCTCCTGGAAAATGAGGATG 1553

Qy       3839 AACAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAG 3898
              | ||||| || || |||||||||||| ||||||||| | |||||||||||||||||||  |
Db       1554 AGCAGAACAGCCCACCAAAAAAGGGCAAAAGAGGCAGGCCACCAAAACCTCTTGGTGGGG 1613

Qy       3899 GTACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTG 3958
               | ||| | || |||||||| ||||||||||||| || |||||| ||| |||||| ||
Db       1614 GGACATCGAAGGAAGAGCCCACAATGAAAACATCCAAGAAAGGAAACAAGAAAAAAACTTG 1673

Qy       3959 GACCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGA 4018
               |||||| |   ||| || || ||||||||||||||||||| |||||| || ||||| |
Db       1674 TACCTCCTGTAGTAGACGACGATGAAGAAGAAGAAAGACAAATTGGAAACACAGAACATA 1733

Qy       4019 AGTCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTG 4078
              |||| |||||||||||||||||       |||| ||||||||| ||||||||||||||||||
Db       1734 AGTCAAAAAGCAAACAGCACCGCACATCAAAGAGAGCACAACAGAGAGCAGAATCTCCTG 1793

Qy       4079 AATCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAA 4138
              || |  ||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db       1794 AAACAAGTGCAGTTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAA 1853

Qy       4139 AAACGCCATCACCATCACAACCAAAAAAAAA 4169
              || |  ||||||||||||||||     |||||
Db       1854 AAGCACCATCACCATCACAACCGCCCAAAAA 1884
```

RESULT 5
BC032988

```
LOCUS       BC032988                1359 bp    mRNA    linear   HTC 20-SEP-2002
DEFINITION  Mus musculus, clone IMAGE:1179696, mRNA.
ACCESSION   BC032988
VERSION     BC032988.1  GI:21426933
KEYWORDS    HTC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1359)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Marcello Bento Soares, Ph.D.
            cDNA Library Preparation: Soares Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 66 Row: c Column: 13
            This clone has the following problem: no polyA-tail.
FEATURES             Location/Qualifiers
     source          1..1359
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1179696"
                     /tissue_type="Mammary gland"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac"
BASE COUNT      505 a    296 c    304 g    254 t
ORIGIN

  Query Match             26.2%;  Score 1095;  DB 11;  Length 1359;
  Best Local Similarity   88.5%;  Pred. No. 1e-177;
  Matches 1201;  Conservative    0;  Mismatches  150;  Indels    6;  Gaps    1;

Qy       2813 AGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATG 2872
              |||| ||||| |||||||| ||||||||||| ||||| ||||| || |||||  |||| ||||
Db          1 AGTACATGGCCATCTGTGCTCTTTGTGCCAAAGACCCTGTGAAAGAGAGGCGAGCCCATG 60

Qy       2873 CTAGGCAATGTTTGGTGAAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAG 2932
              |||| || || |||||||| || || | ||| ||| ||||||| |||||||||||||||
Db         61 CTAGACAGTGCTTGGTGAAGAACATCACTGTGAGGAGGGAGTACCTGAAGCAGCATGCAG 120

Qy       2933 CTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACC 2992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACC 180

Qy       2993 TTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTA 3052
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        181 TTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTGA 240

Qy       3053 AAGAATGTCTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACG 3112
              |||||| ||||||||||||||||| ||||| |||||||||||||||||| |||| | |
Db        241 AAGAATGCCTTTGGTTTGTTCTGGAGATATTGATGGCTAAAAATGAAAACAACAGCCATG 300

Qy       3113 CTTTTATCAGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATG 3172
              | ||||||||||| |||||||||||||||||||| ||||||||||| ||||||||||||
Db        301 CATTTATCAGAAAAATGGTAGAAAATATTAAACAGACAAAAGATGCTCAAGGACCAGATG 360
```

```
Qy       3173 ATGCAAAAATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGT 3232
Db        361 ATACAAAAATGAATGAAAAATTGTACACCGTGTGTGATGTTGCCATGAACATCATCATGT 420

Qy       3233 CAAAGAGTACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCT 3292
Db        421 CCAAGAGCACCACGTACAGCCTGGAGTCTCCTAAGGACCCCGTGCTGCCAGCTCGGTTTT 480

Qy       3293 TCACTCAACCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAAT 3352
Db        481 TCACCCAGCCTGACAAGAATTTTAGTAACACCAAAAATTACCTGCCTCCAGAAATGAAAT 540

Qy       3353 CATTTTTCACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCAC 3412
Db        541 CATTTTTCACTCCTGGAAAACCTAAAACAGCCAATGTTCTCGGAGCAGTTAATAAGCCAC 600

Qy       3413 TTTCATCAGCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATG 3472
Db        601 TTTCATCAGCAGGCAAACAGTCTCAGACCAAATCATCAAGAATGGAAACTGTGAGCAACG 660

Qy       3473 CAAGCAGCAGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTG 3532
Db        661 CAAGCAGCAGCTCCAACCCAAGCTCTCCTGGAAGGATCAAGGGGAGGCTTGATAGCTCTG 720

Qy       3533 AAATGGATCACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAA 3592
Db        721 AAATGGATCACAGTGAAAATGAAGATTATACAATGTCTTCACCTTTGCCAGGAAAAAAAA 780

Qy       3593 GTGACAAGAGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGA 3652
Db        781 GTGACAAGAGAGAAGACCCTGATCTT------TCTGAGTTGGAGAAGCCTAGAAGTCGGA 834

Qy       3653 AAAAAACGCCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGG 3712
Db        835 AAAAAGCACCTGTCACAGACCCTGAAGAGAAATTAGGTATGGATGACCTAACTAAGTTGG 894

Qy       3713 TACAGGAACAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAG 3772
Db        895 TACAGGAACAGAAACCTAAAGGCAGTCAGCGAGGCCGGAAAAGAGGCCGTACAGCTTCAG 954

Qy       3773 AATCTGATGAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATG 3832
Db        955 ACTCAGACGAGCAGCAGTGGCCTGAGGAGAAGAGGCACAAAGAGGAGCTCCTGGAAAATG 1014

Qy       3833 AAGATGAACAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTG 3892
Db       1015 AGGATGAGCAGAACAGCCCACCAAAAAAGGGCAAAAGAGGCAGGCCACCAAAACCTCTTG 1074

Qy       3893 GTGGAGGTACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAA 3952
Db       1075 GTGGGGGACATCGAAGGAAGAGCCCACAATGAAAACATCCAAGAAAGGAAACAAGAAAA 1134

Qy       3953 AATCTGGACCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGG 4012
Db       1135 AACTTGTACCTCCTGTAGTAGACGACGATGAAGAAGAAGAAAGACAAATTGGAAACACAG 1194

Qy       4013 AACAGAAGTCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAAT 4072
Db       1195 AACATAAGTCAAAAAGCAAACAGCACCGCACATCAAAGAGAGCACAACAGAGAGCAGAAT 1254

Qy       4073 CTCCTGAATCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGAC 4132
Db       1255 CTCCTGAAACAAGTGCAGTTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGAC 1314

Qy       4133 CATCAAAAACGCCATCACCATCACAACCAAAAAAAAA 4169
Db       1315 CATCAAAAGCACCATCACCATCACAACCCCCCAAAAA 1351
```

```
RESULT 6
BF509252
LOCUS       BF509252                 791 bp    mRNA    linear   EST 06-DEC-2000
DEFINITION  UI-H-BI4-aow-c-07-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone
            IMAGE:3086220 3', mRNA sequence.
ACCESSION   BF509252
VERSION     BF509252.1  GI:11592550
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 791)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            NCI-CGAP clone distribution information can be found through the
            I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .791
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3086220"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_Sub8"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub8
                     is a subtracted library derived from NCI_CGAP_Sub5. The
                     NCI_CGAP_Sub8 library had 2.5 million recombinants. A
                     single-stranded DNA preparation of NCI_CGAP_Sub5 was used
                     as a tracer in a subtractive  hybridization with a driver
                     comprising: a pool of clones from  NCI_CGAP_Sub5 (IMAGE
                     clone Ids 2732833-2737415, 3068040-3069191; 25% of the
                     driver population), a pool of clones from NCI_CGAP_Sub4
                     (IMAGE clone Ids 2723592-2729326; 25% of the driver
                     population), NCI_CGAP_Sub6 (pool AIF-AJU, IMAGE Ids
                     2728969-2733190; 25% of the driver population), and
                     NCI_CGAP_Sub7 (IMAGE Ids 3069192-3072238,  3081864-3084550
                     ; 25% of the driver population).  Subtraction was
                     performed as previously described [Bonaldo, Lennon &
                     Soares (1996): Normalization and Subtraction: Two
                     Approaches To Facilitate Gene Discovery. Genome Research
                     6, 791-806.
                     TAG_LIB=NCI_CGAP_Lu5
                     TAG_TISSUE=lung
                     TAG_SEQ=CAAC"
BASE COUNT      270 a    151 c    144 g    226 t
ORIGIN

  Query Match             17.2%;  Score 716;  DB 10;  Length 791;
  Best Local Similarity   99.6%;  Pred. No. 1e-112;
  Matches  749;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    3;

Qy       2680 TGTTACCATGAAATCATCACATTAGAACAATATCAGCTATGTGCATTAGCTATCAACGAT 2739
Db         13 TGTTACCATGAAATCATCACATTAGAACAATATCAGCTATGTGCATTAGCTATCAACGAT 72

Qy       2740 GAATGCTATCAAGTAAGACAAGTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTA 2799
Db         73 GAATGCTATCAAGTAAGACAAGTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTA 132

Qy       2800 CGGCTTCCACTTGAGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAG 2859
Db        133 CGGCTTCCACTTGAGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAG 192

Qy       2860 AGAAGAGCTCATGCTAGGCAATGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTG 2919
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         193 AGAAGAGCTCATGCTAGGCAATGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTG 252

Qy        2920 AAGCAGCATGCAGCTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCA 2979
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         253 AAGCAGCATGCAGCTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCA 312

Qy        2980 TATACAATTCACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAA 3039
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         313 TATACAATTCACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAA 372

Qy        3040 CTTAAAGATGTTAAAGAATGTCTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAA 3099
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         373 CTTAAAGATGTTAAAGAATGTCTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAA 432

Qy        3100 AATAACAGTCACGCTTTTATCAGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCC 3159
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         433 AATAACAGTCACGCTTTTATCAGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCC 492

Qy        3160 CAAGGACCAGATGATGCAAAAATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATG 3219
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         493 CAAGGACCAGATGATGCAAAAATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATG 552

Qy        3220 AATATCATCATGTCAAAGAGTACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTA 3279
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         553 AATATCATCATGTCAAAGAGTACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTA 612

Qy        3280 CCAGCTCGTTTCTTCACTCAACCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCT 3339
               |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db         613 CCAGCTCGTTTCTTCACTCAACCTGACAAG-ATTTCAGTAACACCAAAAATTATCTGCCT 671

Qy        3340 CCTGAAATGAAATCATTTTTCACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCT 3399
               ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db         672 CCTGAAATGAAATCATTTTTCACTCCTGGAAAACCTAAAACAA-CAATGTTCTAGGAGCT 730

Qy        3400 GTTAACAAGCCACTTTCATCAGCAGGCAAGCA 3431
               ||||||||| ||||||||||||||||||||||
Db         731 GTTAACAAG-CACTTTCATCAGCAGGCAAGCA 761
```

```
RESULT 7
CA980171
LOCUS       CA980171                 926 bp    mRNA    linear   EST 06-JAN-2003
DEFINITION  AGENCOURT_11295822 NIH_MGC_164 Mus musculus cDNA clone
            IMAGE:30145391 5', mRNA sequence.
ACCESSION   CA980171
VERSION     CA980171.1  GI:27512825
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 926)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. David Rowe and Dr. Mina
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM0059  row: f  column: 24
            High quality sequence stop: 620.
FEATURES             Location/Qualifiers
     source          1. .926
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30145391"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_164"
                     /note="Vector: pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2:
                     NotI; Non-normalized full-length enriched library from
                     pooled mouse embryonic limb, maxilla and mandible, day
                     10.5 and 11.5 (size selected for the 0.5-1 kb fragments)
                     Cloned directionally, priming method: Oligo-dT.  cDNA
                     enrichment: >1k bp, Average insert size 1.8k bp. Priming
                     sequence: 5'GACTAGTTCTAGATCGCGAGCGGCCGCCC(T) 3'. Tissue
                     contributed by, David Rowe. Library constructed by ResGen,
                     Invitrogen Corp."
BASE COUNT      332 a    199 c    222 g    172 t      1 others
ORIGIN

  Query Match             16.6%;  Score 691.6;  DB 14;  Length 926;
  Best Local Similarity   87.6%;  Pred. No. 1.5e-108;
  Matches  790;  Conservative    0;  Mismatches  109;  Indels    3;  Gaps    3;

Qy        3042 TAAAGATGTTAAAGAATGTCTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAA 3101
               ||||||||| |||||||| ||||||||||||||||| ||||| |||||||||||||||||
Db          25 TAAAGATGTGAAAGAATGCCTTTGGTTTGTTCTGGAGATATTGATGGCTAAAAATGAAAA 84

Qy        3102 TAACAGTCACGCTTTTATCAGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCA 3161
                ||||| || || ||||||||||| |||||||||||||||||||| ||||||||||| ||
Db          85 CAACAGCCATGCATTTATCAGAAAAATGGTAGAAAATATTAAACAGACAAAAGATGCTCA 144

Qy        3162 AGGACCAGATGATGCAAAAATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAA 3221
               ||||||||||||| ||||||||||||||||| ||||||| ||||||||||||||||||||
Db         145 AGGACCAGATGATACAAAAATGAATGAAAAATTGTACACCGTGTGTGATGTTGCCATGAA 204

Qy        3222 TATCATCATGTCAAAGAGTACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACC 3281
                ||||||||||| ||||| || || |||||  |||| |||||||| ||||| || || ||
Db         205 CATCATCATGTCCAAGAGCACCACGTACAGCCTGGAGTCTCCTAAGGACCCCGTGCTGCC 264

Qy        3282 AGCTCGTTTCTTCACTCAACCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCC 3341
               |||||| || ||||| || |||||||||||||| ||||||||||||||||| ||||||||
Db         265 AGCTCGGTTTTTCACCCAGCCTGACAAGAATTTTAGTAACACCAAAAATTACCTGCCTCC 324

Qy        3342 TGAAATGAAATCATTTTTCACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGT 3401
                ||||||||||||||||||||||||||||||||||||||| |||||||||| ||||| ||
Db         325 AGAAATGAAATCATTTTTCACTCCTGGAAAACCTAAAACAGCCAATGTTCTCGGAGCAGT 384

Qy        3402 TAACAAGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAAC 3461
               ||| ||||||||||||||||||||||| || |||||||||||||||||| ||||||||||
Db         385 TAATAAGCCACTTTCATCAGCAGGCAAACAGTCTCAGACCAAATCATCAAGAATGGAAAC 444

Qy        3462 TGTAAGCAATGCAAGCAGCAGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCT 3521
               ||| ||||| |||||||||||||| || ||||||||||||||||| || |||||||||||
Db         445 TGTGAGCAACGCAAGCAGCAGCTCCAACCCAAGCTCTCCTGGAAGGATCAAGGGGAGGCT 504

Qy        3522 TGATAGTTCTGAAATGGATCACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCC 3581
               |||||| |||||||||||||||||||||||||||||||| ||  |||||||||||||||||
Db         505 TGATAGCTCTGAAATGGATCACAGTGAAAATGAAGATTATAC-ATGTCTTCACCTTTGCC 563

Qy        3582 GGGGAAAAAAAGTGACAAGAGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCC 3641
                || |||||||||||||||||||| ||| |||||||||||||||||||  ||||||||||
Db         564 AGGAAAAAAAAGTGACAAGAGAGAAGACCCTGATCTTGTAAGGTCTGAGCTGGAGAAGCC 623

Qy        3642 TAGAGGCAGGAAAAAAACGCCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTT 3701
               |||| |  |||||| | || | ||||||| |  || ||||||||||||||||||||| |
Db         624 TAGAAGTCGGAAAAAAGCACCTGTCACAGACCCTGAAGAGAAATTAGGTATGGATGACCT 683

Qy        3702 GACTAAGTTGGTACAGGAACAGAAACCTAAAGGCAGTCAGCGAAGTCGG-AAAAGAGGCC 3760
                |||||||||||||||||||||||||||||||||||||||||| | ||| ||||||||||
Db         684 AACTAAGTTGGTACAGGAACAGAAACCTAAAGGCAGTCAGCGAGGCCGGAAAAAGAGGCC 743

Qy        3761 ATACGGCTTCAGAATCTGATGAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATA 3820
                ||| ||||||| || || || ||||||||||||||||| ||||||| |||||| ||
Db         744 GTACAGCTTCAGACTCAGACGAGCAGCAGTGGCCTGAGGAGAAGAGGCACAAAGAGGAGC 803

Qy        3821 TATTAGAAAATGAAGATGAACAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCAC 3880
               |  | |||||||| ||||| |||||    || ||||||||||| ||||||||| | ||
Db         804 TCCTGGAAAATGAGGATGAGCAGAAACAGCCCCCAAAAAAGGGCAAAAGAGGCAGGCCCA 863
```

```
Qy       3881 CAAAACCTCTTGG-TGGAGGTACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAA 3939
              ||||||||||||| ||| || ||| | || | ||| || || |    ||||||| || |||
Db        864 CAAAACCTCTTGGNTGGGGGACATCGAAGGGAGAACCCACCAATGAAACTTCCAAGAAA 923

Qy       3940 GG 3941
               |
Db        924 AG 925


RESULT 8
BU940978
LOCUS       BU940978                 745 bp    mRNA    linear   EST 18-OCT-2002
DEFINITION  AGENCOURT_10538461 NIH_MGC_128 Homo sapiens cDNA clone
            IMAGE:6712336 5', mRNA sequence.
ACCESSION   BU940978
VERSION     BU940978.1  GI:24129797
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 745)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: NCI
             cDNA Library Preparation: Michael Brownstein Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM3020  row: n  column: 16
            High quality sequence stop: 553.
FEATURES             Location/Qualifiers
     source          1..745
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6712336"
                     /tissue_type="mixed (pool of 40 RNAs)"
                     /lab_host="DH10B (T1-phage-resistant)"
                     /clone_lib="NIH_MGC_128"
                     /note="Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc);
                     Site_2: SfiI (ggccgcctcggcc); Double-stranded cDNA was
                     prepared from a pool of 40 cell line polyA+ RNAs (bladder
                     - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -
                     4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
                     kidnney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%
                     , ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary
                     gland - 1.3%, and skin - 2.3%).  5' and 3' adaptors were
                     used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'.  Full-length
                     enriched library was constructed using the Clontech
                     Creator SMART kit and size-selected to contain the >2 kb
                     size fraction (other fractions present in NIH_MGC_126 and
                     NIH_MGC_127).  Library created in the laboratory of T.
                     Usdin, M.D., Ph.D. (NIMH, NIH). Note: this is a NIH_MGC
                     Library."
BASE COUNT      252 a    142 c    152 g    199 t
ORIGIN

  Query Match             16.4%;  Score 685.8;  DB 13;  Length 745;
  Best Local Similarity   99.6%;  Pred. No. 1.5e-107;
  Matches  698;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1414 GAACGGATGAAATGCTTATATTACTTGTATGCCACACTGGATTTAAATGCTGTGAAAGCA 1473
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 GAACGGATGAAATGCTTATATTACTTGTATGCCACACTGGATTTAAATGCTGTGAAAGCA 62

Qy       1474 TTGAATGAAATGTGGAAATGTCAAAATCTGCTCCGACATCAAGTAAAGGATTTGCTTGAC 1533
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 TTGAATGAAATGTGGAAATGTCAAAATCTGCTCCGACATCAAGTAAAGGATTTGCTTGAC 122

Qy       1534 TTGATTAAGCAACCCAAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGGTT 1593
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 TTGATTAAGCAACCCAAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGGTT 182

Qy       1594 ATTACAAGAAATTTACCTGATCCTGGTAAGGCTCAGGATTTCATGAAGAAATTCACACAG 1653
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 ATTACAAGAAATTTACCTGATCCTGGTAAGGCTCAGGATTTCATGAAGAAATTCACACAG 242

Qy       1654 GTGTTAGAAGATGATGAGAAAATAAGAAAGCAGTTAGAAGTACTTGTTAGTCCAACATGC 1713
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 GTGTTAGAAGATGATGAGAAAATAAGAAAGCAGTTAGAAGTACTTGTTAGTCCAACATGC 302

Qy       1714 TCCTGCAAGCAGGCTGAAGGTTGTGTGCGTGAAATAACTAAGAAGTTGGGCAACCCCAAA 1773
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 TCCTGCAAGCAGGCTGAAGGTTGTGTGCGTGAAATAACTAAGAAGTTGGGCAACCCCAAA 362

Qy       1774 CAGCCTACAAATCCTTTCCTGGAAATGATCAAGTTTCTCTTGGAGAGGATAGCACCTGTG 1833
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 CAGCCTACAAATCCTTTCCTGGAAATGATCAAGTTTCTCTTGGAGAGGATAGCACCTGTG 422

Qy       1834 CACATAGATACCGAATCTATCAGTGCTCTTATTAAACAAGTGAACAAATCAATAGATGGA 1893
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 CACATAGATACCGAATCTATCAGTGCTCTTATTAAACAAGTGAACAAATCAATAGATGGA 482

Qy       1894 ACAGCAGATGATGAAGATGAGGGTGTTCCAACTGATCAAGCCATCAGAGCAGGTCTTGAA 1953
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 ACAGCAGATGATGAAGATGAGGGTGTTCCAACTGATCAAGCCATCAGAGCAGGTCTTGAA 542

Qy       1954 CTGCTTAAGGTACTCTCATTTACACATCCCATCTCATTTCATTCTGCTGAAACATTTGAA 2013
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 CTGCTTAAGGTACTCTCATTTACACATCCCATCTCATTTCATTCTGCTGAAACATTTGAA 602

Qy       2014 TCATTACTGGCTTGTCTGAAAATGGATGATGAAAAAGTAGCAGAAGCTGCACTACAAATT 2073
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 TCATTACTGGCTTGTCTGAAAATGGATGATGAAAAAGTAGCAGAAGCTGCACTACAAATT 662

Qy       2074 TTCAAAAACACAGGAAG-CAAAATTGAAGAGGATTTTCCAC 2113
              ||||||||| ||||||| ||||||||||||||||||| |||
Db        663 TTCAAAAACCCAGGAAGCCAAAATTGAAGAGGATTTTTCAC 703


RESULT 9
BM794657
LOCUS       BM794657                 692 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  K-EST0076048 S22SNU16n1 Homo sapiens cDNA clone S22SNU16n1-50-A08
            5', mRNA sequence.
ACCESSION   BM794657
VERSION     BM794657.1  GI:19142889
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 692)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 50  row: A  column: 08
            High quality sequence stop: 692.
```

```
FEATURES             Location/Qualifiers
     source          1. .692
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="S22SNU16n1-50-A08"
                     /sex="F"
                     /tissue_type="Ascites"
                     /cell_type="Lymphoblast-like"
                     /cell_line="SNU-16"
                     /lab_host="DH10B"
                     /clone_lib="S22SNU16n1"
                     /note="Organ: Stomach; Vector: pT7T3-Pac; Site_1: EcoRI;
                     Site_2: NotI; The S22SNU16 library was contributed by the
                     Soares laboratory and it was constructed as described by
                     Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome
                     Research 6(9): 791-806. RNA was prepared from harvested
                     cells of SNU-16 culture. SNU-16 cell was obtained from
                     Korean Cell Line Bank (KCLB). SNU-16 was established from
                     ascitic fluids of Korean patients by Park J.G. et al.
                     (1990), Cancer Res 50: 2773-2780."
BASE COUNT      264 a    136 c    148 g    144 t
ORIGIN

  Query Match             16.2%;  Score 677;  DB 12;  Length 692;
  Best Local Similarity   99.3%;  Pred. No. 4.9e-106;
  Matches  680;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       3186 TGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGTACTAC 3245
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGTACTAC 60

Qy       3246 ATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAACCTGA 3305
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAACCTGA 120

Qy       3306 CAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTCACTCC 3365
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTCACTCC 180

Qy       3366 TGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCAGCAGG 3425
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCAGCAGG 240

Qy       3426 CAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGCAGCTC 3485
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGCAGCTC 300

Qy       3486 AAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGATCACAG 3545
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGATCACAG 360

Qy       3546 TGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAGAGAGA 3605
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAGAGAGA 420

Qy       3606 CGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACGCCCGT 3665
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACGCCCGT 480

Qy       3666 CACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAACAGAA 3725
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAACAGAA 540

Qy       3726 ACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGATGAACA 3785
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGATGAACA 600

Qy       3786 GCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAA 3845
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAA 660

Qy       3846 TAGTCCGCCAAAAAAGGGTAAAAGA 3870
              |||||||||||||||    |||| |
Db        661 TAGTCCGCCAAAAAAAAAAAAAAAA 685

RESULT 10
CA324097
LOCUS       CA324097                 786 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  UI-M-FY0-cco-m-16-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
            IMAGE: 6822425 5', mRNA sequence.
ACCESSION   CA324097
VERSION     CA324097.1  GI:24542195
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 786)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .786
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6822425"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FY0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      256 a    171 c    165 g    191 t      3 others
ORIGIN

  Query Match             16.1%;  Score 671;  DB 14;  Length 786;
  Best Local Similarity   90.7%;  Pred. No. 5.3e-105;
  Matches  713;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy       2794 CGTTTACGGCTTCCACTTGAGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTA 2853
              || ||||||||||| |||||||| ||||| |||||||| |||||||| ||||| ||||| 
Db          1 CGCTTACGGCTTCCCCTTGAGTACATGGCCATCTGTGCTCTTTGTGCCAAAGACCCTGTG 60

Qy       2854 AAGGAGAGAAGAGCTCATGCTAGGCAATGTTTGGTGAAAAATATAAATGTAAGGCGGGAG 2913
              || |||||  |||| |||||||| || || |||||||| || || | ||| ||| |||||
Db         61 AAAGAGAGGCGAGCCCATGCTAGACAGTGCTTGGTGAAGAACATCACTGTGAGGAGGGAG 120
```

```
Qy       2914 TATCTGAAGCAGCATGCAGCTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATGTT 2973
              || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TACCTGAAGCAGCATGCAGCTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATGTT 180

Qy       2974 GTTCCATATACAATTCACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATATT 3033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTTCCATATACAATTCACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATATT 240

Qy       3034 GAACAACTTAAAGATGTTAAAGAATGTCTTTGGTTTGTTCTGGAAATATTAATGGCTAAA 3093
              ||||||||||||||||| |||||||| ||||||||||||||||| ||||| |||||||||
Db        241 GAACAACTTAAAGATGTGAAAGAATGCCTTTGGTTTGTTCTGGAGATATTGATGGCTAAA 300

Qy       3094 AATGAAAATAACAGTCACGCTTTTATCAGAAAGATGGTAGAAAATATTAAACAAACAAAA 3153
              |||||||| ||||| || || ||||||||||| |||||||||||||||||||| ||||||
Db        301 AATGAAAACAACAGCCATGCATTTATCAGAAAAATGGTAGAAAATATTAAACAGACAAAA 360

Qy       3154 GATGCCCAAGGACCAGATGATGCAAAAATGAATGAAAAACTGTACACTGTGTGTGATGTT 3213
              ||||| ||||||||||||||| ||||||||||||||||| ||||||| ||||||||||||
Db        361 GATGCTCAAGGACCAGATGATACAAAAATGAATGAAAAATTGTACACCGTGTGTGATGTT 420

Qy       3214 GCCATGAATATCATCATGTCAAAGAGTACTACATACAGTTTGGAATCTCCTAAAGACCCG 3273
              |||||||| ||||||||||| ||||| || ||| ||||  |||| |||||||| ||||| 
Db        421 GCCATGAACATCATCATGTCCAAGAGCACCACGTACAGCCTGGAGTCTCCTAAGGACCCC 480

Qy       3274 GTACTACCAGCTCGTTTCTTCACTCAACCTGACAAGAATTTCAGTAACACCAAAAATTAT 3333
              || || |||||||| || ||||| || |||||||||||||| ||||||||||||||||| 
Db        481 GTGCTGCCAGCTCGGTTTTTCACCCAGCCTGACAAGAATTTTAGTAACACCAAAAATTAC 540

Qy       3334 CTGCCTCCTGAAATGAAATCATTTTCACTCCTGGAAAACCTAAAACAACCAATGTTCTA 3393
              |||||||| |||||||||||||||||||||||||||||||||||||| ||||||||||
Db        541 CTGCCTCCAGAAATGAAATCATTTTCACTCCTGGAAAACCTAAAACAGCCAATGTTCTC 600

Qy       3394 GGAGCTGTTAACAAGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAAATCATCACGA 3453
              ||||| ||||| ||||||||||||||||||||||| || |||||||||| ||||||| ||
Db        601 GGAGCAGTTAATAAGCCACTTTCATCAGCAGGCAAACAGTCTCAGACCANATCATCAAGA 660

Qy       3454 ATGGAAACTGTAAGCAATGCAAGCAGCAGCTCAAATCCAAGCTCTCCTGGAAGAATAAAG 3513
              ||||||||||| ||||| |||||||||||||| |  ||| ||||||||||||| || |||
Db        661 ATGGAAACTGTGAGCAACGCAAGCAGCAGCTCCANCCCANGCTCTCCTGGAAGGATCAAG 720

Qy       3514 GGGAGGCTTGATAGTTCTGAAATGGATCACAGTGAAAATGAAGATTACACAATGTCTTCA 3573
              |||||||||||||| |||||||||||||||||||||||||||||||| ||||||||||||
Db        721 GGGAGGCTTGATAGCTCTGAAATGGATCACAGTGAAAATGAAGATTATACAATGTCTTCA 780

Qy       3574 CCTTTG 3579
              ||||||
Db        781 CCTTTG 786
```

RESULT 11
BG256731

```
LOCUS       BG256731                 734 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  602370939F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4479153 5',
            mRNA sequence.
ACCESSION   BG256731
VERSION     BG256731.1  GI:12766547
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 734)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10310  row: m  column: 10
            High quality sequence stop: 732.
FEATURES             Location/Qualifiers
     source          1..734
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4479153"
                     /tissue_type="embryonal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_92"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      278 a     154 c     161 g     141 t
ORIGIN

  Query Match             15.9%;  Score 663.4;  DB 10;  Length 734;
  Best Local Similarity   96.6%;  Pred. No. 1.1e-103;
  Matches  700;  Conservative    0;  Mismatches   21;  Indels    4;  Gaps    2;

Qy       3232 TCAAAGAGTACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTC 3291
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db          1 TCAAAGAGTACTACATACAG-TTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTC 59

Qy       3292 TTCACTCAACCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAA 3351
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 TTCACTCAACCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAA 119

Qy       3352 TCATTTTTCACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCA 3411
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 TCATTTTTCACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCA 179

Qy       3412 CTTTCATCAGCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAAT 3471
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 CTTTCATCAGCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAAT 239

Qy       3472 GCAAGCAGCAGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCT 3531
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 GCAAGCAGCAGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCT 299

Qy       3532 GAAATGGATCACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAA 3591
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 GAAATGGATCACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAA 359

Qy       3592 AGTGACAAGAGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGG 3651
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 AGTGACAAGAGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGG 419

Qy       3652 AAAAAAACGCCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTG 3711
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 AAAAAAACGCCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTG 479

Qy       3712 GTACAGGAACAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCA 3771
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 GTACAGGAACAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCA 539

Qy       3772 GAATCTGATGAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAA- 3830
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| | 
Db        540 GAATCTGATGAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAACAA 599

Qy       3831 --TGAAGATGAACAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCT 3888
                  | |||||||||||||||||||||||| | ||  | |   | |   |  ||||||||
Db        600 CTGAACGATGAACAGAATAGTCCGCCAAAACAAGGGTACAAGAGGCCGACCACAAAACCT 659

Qy       3889 CTTGGTGGAGGTACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAGGAAGCAAA 3948
              |||||||||||||||||||||||||||||||| ||||||||||||| |||||||||||||
Db        660 CTTGGTGGAGGTACACCAAAAGAAGAGCCAACGATGAAAACTTCTATAAAGGAAGCAAA 719

Qy       3949 AAAAA 3953
```

```
                 ||||
Db          720 CAAAA 724

RESULT 12
CB312535
LOCUS       CB312535                 771 bp    mRNA    linear   EST 04-MAR-2003
DEFINITION  AGENCOURT_11835698 NICHD_Rh_Ov1 Macaca mulatta cDNA clone
            IMAGE:6892555 5', mRNA sequence.
ACCESSION   CB312535
VERSION     CB312535.1  GI:28835251
KEYWORDS    EST.
SOURCE      Macaca mulatta (rhesus monkey)
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 771)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Eliot Spindel
             cDNA Library Preparation: CLONTECH
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM3155  row: g  column: 18
            High quality sequence stop: 563.
FEATURES             Location/Qualifiers
     source          1. .771
                     /organism="Macaca mulatta"
                     /mol_type="mRNA"
                     /db_xref="taxon:9544"
                     /clone="IMAGE:6892555"
                     /tissue_type="Ovary"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NICHD_Rh_Ov1"
                     /note="Organ: ovary; Vector: pDNR-LIB; Site_1: Sfi I;
                     Site_2: Sfi I; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.0-4.0 kb. Tissue pooled from
                     pre-pubertal, post pubertal sn menopausal monkeys.
                     Constructed by Clontech. Note: this is a NICHD Library."
BASE COUNT      291 a    144 c    159 g    177 t
ORIGIN

  Query Match            15.8%;  Score 660.6;  DB 14;  Length 771;
  Best Local Similarity  96.0%;  Pred. No. 3.2e-103;
  Matches  744;  Conservative    0;  Mismatches   19;  Indels   12;  Gaps    6;

Qy       3014 ATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGTCTTTGGTTTGTTC 3073
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 ATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGTCTTTGGTTTGTTC 62

Qy       3074 TGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATCAGAAAGATGGTAG 3133
              ||||||||||||||||||||||||||||||||||||| || |||||||||||||||||||
Db         63 TGGAAATATTAATGGCTAAAAATGAAAATAACAGTCATGCATTTATCAGAAAGATGGTAG 122

Qy       3134 AAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAAATGAATGAAAAAC 3193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 AAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAAATGAATGAAAAAC 182

Qy       3194 TGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGTACTACATACAGTT 3253
              |||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 TGTATACCGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGTACTACATACAGTT 242

Qy       3254 TGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAACCTGACAAGAATT 3313
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        243 TGGAATCTCCTAAAGACCCAGTACTACCAGCTCGTTTCTTCACTCAACCTGACAAGAATT 302

Qy       3314 TCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTCACTCCTGGAAAAC 3373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 TCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTCACTCCTGGAAAAC 362

Qy       3374 CTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCAGCAGGCAAGCAAT 3433
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||| ||||
Db        363 CTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCGCTTTCATCAGCAGGCAAACAAT 422

Qy       3434 CTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGCAGCTCAAATCCAA 3493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 CTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGCAGCTCAAATCCAA 482

Qy       3494 GCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGATCACAGTGAAAATG 3553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 GCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGATCACAGTGAAAATG 542

Qy       3554 AAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAGAGAGACGACTCTG 3613
              ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db        543 AAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAACGAGACGACTCTG 602

Qy       3614 ATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACGCCCGTCACAGAAC 3673
              |||||      ||||||||||||||||||||||||||||||||||| |||||||||||||
Db        603 ATCTT------TCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACACCCGTCACAGAAC 656

Qy       3674 AGGAG-GAGAAATTAGGTATGGATGAC-TTGACTAAGTTGGTACAGGAAC-AGAAACCTA 3730
              ||||| ||||||||||||||||||||| |||||||||||||||||||||| |||||||||
Db        657 AGGAGAGAGAAATTAGGTATGGATGACTTTGACTAAGTTGGTACAGGAACAAGAAACCTA 716

Qy       3731 AAGGCAGTCAG-CGAAGTCGGAAAAGA--GGCCATACGGCTTCAGAATCTGATGA 3782
              ||||||||||| ||| ||||||||||  | |      |||||  |||||||||||
Db        717 AAGGCAGTCAGCCGAGGTCGGAAAAGAAGGCCATACGGCTTTCGAATCTGATGA 771

RESULT 13
CA405800
LOCUS       CA405800                 695 bp    mRNA    linear   EST 07-NOV-2002
DEFINITION  1001949 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens
            cDNA 5', mRNA sequence.
ACCESSION   CA405800
VERSION     CA405800.1  GI:24770671
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 695)
  AUTHORS   Yang,R.-Z., Shuldiner,A. and Gong,D.-W.
  TITLE     EST analysis of human adipose gene expression
  JOURNAL   Unpublished
COMMENT     Contact: Gong Da-Wei
            Division of Endocrinology, Diabetes and Nutrition
            University of Maryland
            660 Redwood St, HH497, Baltimore, MD 21201, USA
            Tel: 410 706 1672
            Fax: 410 706 1622
            Email: dgong@medicine.umaryland.edu
            PCR PRimers
            FORWARD: CTCGGGAAGCGCGCCATTGTGTTGGT
            BACKWARD: AATACGACTCACTATAGGGCGAATTGG
            Seq primer: GTTGGTACCCGGGAATTC.
FEATURES             Location/Qualifiers
     source          1. .695
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /sex="Male and Female"
                     /tissue_type="Adipose"
                     /clone_lib="Human Fat Cell 5'-Stretch Plus cDNA Library"
                     /note="Vector: lambdaTriplEx"
BASE COUNT      222 a    154 c    128 g    191 t
ORIGIN
```

```
  Query Match             15.7%;  Score 657.2;  DB 14;  Length 695;
  Best Local Similarity   99.1%;  Pred. No. 1.2e-102;
  Matches  692;  Conservative    0;  Mismatches    3;  Indels    3;  Gaps    3;

Qy       1680 AAAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGT 1739
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGT 60

Qy       1740 GCGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAAT 1799
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAAT 120

Qy       1800 GATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGC 1859
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGC 180

Qy       1860 TCTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGT 1919
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGT 240

Qy       1920 TCCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACA 1979
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACA 300

Qy       1980 TCCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGA 2039
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGA 360

Qy       2040 TGATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGA 2099
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGA 420

Qy       2100 AGAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAA 2159
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAA 480

Qy       2160 AGGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAA 2219
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAA 540

Qy       2220 AGAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCT 2279
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCT 600

Qy       2280 GGAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCA 2339
              ||||||||||||||||||||||||||||||||| ||||||||||||||||| ||||||||
Db        601 GGAACATCTCATAACACCATTGGTTACTATTGG-CATATTGCTCTCCTTGC-CCTGATCA 658

Qy       2340 ATTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCA 2377
              |||||||| |||||| ||||||| |||| |||||||||
Db        659 ATTTGCTG-TCCTTGAAATCTTTGGTAACTACTTTCA 695

RESULT 14
BU703488
LOCUS       BU703488                 786 bp    mRNA    linear   EST 09-OCT-2002
DEFINITION  UI-M-FO0-bzo-g-20-0-UI.r1 NIH_BMAP_FO0 Mus musculus cDNA clone
            IMAGE: 6405163 5', mRNA sequence.
ACCESSION   BU703488
VERSION     BU703488.1  GI:23629371
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 786)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..786
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6405163"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 12.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FO0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is TGAGAGAGCC. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      248 a    143 c    152 g    242 t      1 others
ORIGIN

  Query Match             15.7%;  Score 653.8;  DB 13;  Length 786;
  Best Local Similarity   91.1%;  Pred. No. 4.7e-102;
  Matches  694;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy         19 AGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAGGAAATATCAGATAAAATA 78
              |||||||| ||||| |||||||| || || ||||| ||||||||||| ||||||||||| 
Db         25 AGGACCAACGATGGGAAAATTACTTACCCTCCTGGAGTCAAGGAAATCTCAGATAAAATC 84

Qy         79 TCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAAACTTTTATGGATATGGAC 138
              |||||||||||||||||||| || |||||||||||||| |||||||| ||||| ||||||
Db         85 TCTAAAGAGGAGATGGTGAGGCGGTTAAAGATGGTTGTAAAAACTTTCATGGACATGGAC 144

Qy        139 CAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCTTTACATCTTGCTTCAGAT 198
              |||||||||||||| |||||||| |||||| ||||||||||||||||||||||||| || 
Db        145 CAGGACTCTGAAGAGGAAAAGGAACTTTATCTAAACCTAGCTTTACATCTTGCTTCTGAC 204

Qy        199 TTTTTTCTCAAGCATCCTGGTAAAGATGTTCGCTTACTGGTAGCCTGCTGCCTTGCTGAT 258
              || || ||||||||||||| |||||||||||| |||||||| || |||||||||||||||
Db        205 TTCTTCCTCAAGCATCCTGATAAAGATGTTCGTTTACTGGTGGCTTGCTGCCTTGCTGAT 264

Qy        259 ATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCTGATAAACTAAAGGATATA 318
              ||||||||||||||||||||||| ||||||||||| || || ||||||||||||||||||
Db        265 ATTTTCAGGATTTATGCTCCTGAGGCTCCTTACACGTCTCCCGATAAACTAAAGGATATA 324

Qy        319 TTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACAAAGAGCCCACAATTCAAT 378
              ||||||||||||||||| ||  | ||||| ||||| |||||||||||||| ||||| |||
Db        325 TTTATGTTTATAACAAGGCAACTAAAGGGACTAGAAGATACAAAGAGCCCTCAATTTAAT 384

Qy        379 AGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCATATAACATATGCTTTGAG 438
              ||||||||||||||||||||||||||||| ||||| || |||||||||||||||||||||
Db        385 AGGTATTTTTATTTACTTGAGAACATTGCATGGGTTAAATCATATAACATATGCTTTGAG 444

Qy        439 TTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACCTTATTTTCAGTTATAAAC 498
              |||||||||||||||||||| || || || || |||||||| ||||| ||||||||||||
Db        445 TTAGAAGATAGCAATGAAATCTTTACTCAACTTTACAGAACATTATTCTCAGTTATAAAC 504
```

```
Qy        499 AATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTTATGAGCTCTATTATTTGT 558
Db        505 AATGGCCACAATCAGAAAGTTCATATGCACATGGTGGACCTCATGAGCTCTATCATTTGT 564

Qy        559 GAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTAGTAAATCTGGTACCTGCT 618
Db        565 GAAGGTGATACCGTATCTCAGGAGCTCTTAGATACAGTTTTAGTAAATCTGGTACCTGCC 624

Qy        619 CATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCTTTACTGAAGAGGACAGCT 678
Db        625 CATAAGAACTTAAACAAGCAAGCATATGATTTGGCAAAGGCTTTGCTGAAGAGGACTGCT 684

Qy        679 CAAGCTATTGAGCCATATATTACCACTTTTTTTAATCAGGTTCTGATGCTTGGGAAAACA 738
Db        685 CAAGCTATTGAACCATATATTACCAATTTTTTTAACCAGGTTCTGATGCTTGNGAAAACA 744

Qy        739 TCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 780
Db        745 TCTATCAGTGATTTGTCTGAGCATGTCTTTGATTTGATCTTG 786


RESULT 15
CA324284
LOCUS       CA324284                 779 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  UI-M-FY0-ccp-a-07-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
            IMAGE: 6822512 5', mRNA sequence.
ACCESSION   CA324284
VERSION     CA324284.1  GI:24542382
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 779)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..779
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6822512"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FY0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      241 a     166 c     174 g     197 t      1 others
ORIGIN

  Query Match             15.4%;  Score 640.6;  DB 14;  Length 779;
  Best Local Similarity   89.6%;  Pred. No. 8.6e-100;
  Matches  699;  Conservative    0;  Mismatches   80;  Indels    1;  Gaps    1;

Qy       2585 GAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTTGCTG 2644
Db          1 GGGATTTGACAGAACAAGGGAAAATTAGTAAACCAGATATGTCACGCCTGAGACTTGCTG 60

Qy       2645 CTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACATTAG 2704
Db         61 CTGGGAGTGCTATTGTGAAGCTGGCACAGGAGCCCTGTTACCACGAGATCATTACACTGG 120

Qy       2705 AACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAAGTGT 2764
Db        121 AGCAGTACCAGCTGTGTGCATTAGCCATCAATGATGAGTGTTATCAAGTCAGGCAGGTGT 180

Qy       2765 TTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATGGCAA 2824
Db        181 TCGCTCAGAAACTTCACAAAGGCCTTTCCCGCTTACGGCTTCCCCTTGAGTACATGGCCA 240

Qy       2825 TCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAATGTT 2884
Db        241 TCTGTGCTCTTTGTGCCAAAGACCCTGTGAAAGAGAGGCGAGCCCATGCTAGACAGTGCT 300

Qy       2885 TGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGTGAAA 2944
Db        301 TGGTGAAGAACATCACTGTGAGGAGGGAGTACCTGAAGCAGCATGCAGCTGTTAGTGAAA 360

Qy       2945 AATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCACATG 3004
Db        361 AATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCACATG 420

Qy       3005 ACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGTCTTT 3064
Db        421 ACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTGAAAGAATGCCTTT 480

Qy       3065 GGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATCAGAA 3124
Db        481 GGTTTGTTCTGGAGATATTGATGGCTAAAAATGAAAACAACAGCCATGCATTTATCAGAA 540

Qy       3125 AGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAAATGA 3184
Db        541 AAATGGTAGAAAATATTAAACAGACAAAAGATGCTCAAGGACCAGATGATACAAAAATGA 600

Qy       3185 ATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGTACTA 3244
Db        601 ATGAAAAATTGTACACCGTGTGTGATGTTGCCATGAACATCATCATGTCCAAGAGCACCA 660

Qy       3245 CATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAACCTG 3304
Db        661 CGTACAGCCTGGAGTCTCCTAAGGACCCCGTGCTGCCAGCTCGGTTTTCACCCAGCCTG 720

Qy       3305 ACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTCACTC 3364
Db        721 ACAAGAATTTTAGTAACACCANAAATTACCTGCCTCCAGAAATG-AATCATTTTTCACTC 779


Search completed: September 24, 2003, 16:28:07
Job time : 8082 secs
```

NPTO

OM nucleic - nucleic search, using sw model

Run on: September 24, 2003, 14:13:39 ; Search time 915 Seconds
(without alignments)
11353.711 Million cell updates/sec

Title: US-09-512-581B-3
Perfect score: 4173
Sequence: 1 atggctcattcaaagactag..........cacaaccaaaaaaaaatgtg 4173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1678620 seqs, 1244745471 residues

Total number of hits satisfying chosen parameters: 3357240

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	493	11.8	2496	9	US-09-822-849A-496	Sequence 496, App
	2	411.4	9.9	439	13	US-10-040-739-107	Sequence 107, App
	3	253.8	6.1	1217	9	US-09-925-302-357	Sequence 357, App
c	4	217.6	5.2	471	11	US-09-918-995-3514	Sequence 3514, Ap
c	5	203.6	4.9	530	10	US-09-879-536-682	Sequence 682, App
	6	164	3.9	433	11	US-09-918-995-33817	Sequence 33817, A
	7	96.4	2.3	413	11	US-09-918-995-20720	Sequence 20720, A
	8	87.4	2.1	331	10	US-09-728-445-172	Sequence 172, App
	9	76.8	1.8	642	13	US-10-027-632-105871	Sequence 105871,
c	10	76.8	1.8	739	13	US-10-027-632-19016	Sequence 19016, A
c	11	64.4	1.5	14006	12	US-10-311-455-1931	Sequence 1931, Ap
c	12	62.8	1.5	6668	12	US-10-311-455-1670	Sequence 1670, Ap
	13	61.2	1.5	1267	13	US-10-001-843-45	Sequence 45, Appl
c	14	60.6	1.5	9539	12	US-10-240-453-54	Sequence 54, Appl
c	15	60.6	1.5	9539	14	US-10-239-676-52	Sequence 52, Appl
	16	60	1.4	60	12	US-09-908-975-12584	Sequence 12584, A
c	17	59	1.4	6668	12	US-10-311-455-1669	Sequence 1669, Ap
c	18	58.6	1.4	529	10	US-09-983-965-2109	Sequence 2109, Ap
c	19	58	1.4	446	10	US-09-960-352-3400	Sequence 3400, Ap
c	20	57.8	1.4	16033	12	US-10-311-455-1377	Sequence 1377, Ap
	21	56.8	1.4	196	9	US-09-728-446-1044	Sequence 1044, Ap
c	22	56.6	1.4	17934	12	US-10-311-455-1692	Sequence 1692, Ap
c	23	56.2	1.3	491	10	US-09-878-574-4301	Sequence 4301, Ap
	24	56	1.3	525	14	US-10-198-846-1483	Sequence 1483, Ap
c	25	55.6	1.3	7571	12	US-10-311-455-500	Sequence 500, App
c	26	53.6	1.3	425	10	US-09-960-352-4010	Sequence 4010, Ap
c	27	53.4	1.3	6334	12	US-10-311-455-1186	Sequence 1186, Ap
c	28	53.4	1.3	6485	12	US-10-311-455-1781	Sequence 1781, Ap
	29	53	1.3	1428	13	US-10-001-857-79	Sequence 79, Appl
c	30	53	1.3	16167	12	US-10-311-455-1056	Sequence 1056, Ap
c	31	53	1.3	16167	12	US-10-240-485-82	Sequence 82, Appl
c	32	52.8	1.3	5750	12	US-10-311-455-1981	Sequence 1981, Ap
	33	52.4	1.3	268	10	US-09-960-352-11467	Sequence 11467, A
c	34	52.2	1.3	723	13	US-10-027-632-152119	Sequence 152119,
c	35	52	1.2	6485	12	US-10-311-455-1782	Sequence 1782, Ap
c	36	51.8	1.2	6301	12	US-10-311-455-26	Sequence 26, Appl
c	37	51.8	1.2	11996	12	US-10-240-485-45	Sequence 45, Appl
	38	51.6	1.2	4167	9	US-09-764-878-282	Sequence 282, App
	39	51.6	1.2	4167	9	US-09-764-860-1145	Sequence 1145, Ap
c	40	51.6	1.2	4167	10	US-09-764-846-345	Sequence 345, App
c	41	51.6	1.2	4167	11	US-09-764-891-7710	Sequence 7710, Ap
c	42	51.6	1.2	4167	11	US-09-764-891-10231	Sequence 10231, A
c	43	51.6	1.2	4167	14	US-10-091-483-345	Sequence 345, App
	44	51.6	1.2	4167	14	US-10-079-854-282	Sequence 282, App
	45	51.6	1.2	4167	14	US-10-074-095-1145	Sequence 1145, Ap

ALIGNMENTS

RESULT 1
US-09-822-849A-496
; Sequence 496, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 496
; LENGTH: 2496
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-496

```
  Query Match             11.8%;  Score 493;  DB 9;  Length 2496;
  Best Local Similarity   65.6%;  Pred. No. 2.7e-111;
  Matches  752;  Conservative    0;  Mismatches  390;  Indels    5;  Gaps    2;

Qy       2387 ATCTTCTCATGAATGATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATG 2446
               | |||     || |   |   |      ||||  ||||||      ||||| |||  ||||||||
Db         83 AGCTTGCACAGATTTTGAGTCAACAGGTGAAAAGAATGGAAAACTGTGGTCTCCAGATG 142

Qy       2447 AAGAAGTATCTCCTGAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGC 2506
              |||| || || |||||     | | |    ||    |  ||||| |||||| |    ||||   | ||||
Db        143 AAGAGGTTTCCCCTGAAGTACTAGCAAAGGTACAGGCAATTAAACTTCTGGTAAGGTGGC 202
```

```
Qy       2507 TACTTGGAATGAAAAATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAA 2566
              |  | || ||||||| || ||   |||||| |  | ||| ||| |  | ||  || ||
Db        203 TGTTGGGTATGAAAAACAACCAGTCTAAATCTGCCAATTCAACCCTTCGGTTATTATCAG 262

Qy       2567 CAATATTGCATAGTGATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGT 2626
              | || |||  ||||| || ||| |||||| ||    ||  || |||||| | |||||||
Db        263 CGATGTTGGTTAGTGAGGGTGACCTGACAGAGCAAAAGAGGATCAGTAAATCTGATATGT 322

Qy       2627 CACGTCTGAGACTTGCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACC 2686
              | ||  || || | |||||||| ||||| ||  ||||||| || || ||||| |||||||
Db        323 CTCGCTTGCGATTAGCTGCTGGTAGTGCCATAATGAAGCTTGCTCAGGAACCTTGTTACC 382

Qy       2687 ATGAAATCATCACATTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCT 2746
              |||||| || ||    |||||| | |||||| |||||| | | ||| || ||||| || |
Db        383 ATGAAATTATTACCCCAGAACAGTTTCAGCTCTGTGCACTTGTTATTAATGATGAGTGTT 442

Qy       2747 ATCAAGTAAGACAAGTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTC 2806
              | |||||||| ||  | ||||| ||||| || || || |  |||        |||| ||| |
Db        443 ACCAAGTAAGGCAGATATTTGCTCAGAAGCTGCATAAGGCACTTGTGAAGTTACTGCTCC 502

Qy       2807 CACTTGAGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAG 2866
              || | ||||||||||| |||| |||| | ||||| |||||||||| ||||||||||||
Db        503 CATTGGAGTATATGGCGATCTTTGCCTTGTGTGCCAAAGATCCTGTGAAGGAGAGAAGAG 562

Qy       2867 CTCATGCTAGGCAATGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGC 2926
              | || ||   | ||||||||  ||||||||| | | || |   |||| ||   | ||||||
Db        563 CACACGCACGACAATGTTTACTGAAAAATATCAGTATACGCAGGGAATACATTAAGCAGA 622

Qy       2927 ATGCAGCTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAA 2986
              || |    | || |||  ||||||| || |  | ||  |  |||||| |||||||| |   |
Db        623 ATCCTATGGCTACTGAGAAATTATTATCACTGTTGCCTGAATATGTAGTTCCATACATGA 682

Qy       2987 TTCACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAG 3046
              ||||||   | || ||||| ||||||| |    | |  ||| ||| |||| || |||    |
Db        683 TTCACCTGCTAGCCCATGATCCAGATTTTACAAGATCACAAGATGTTGATCAGCTTCGTG 742

Qy       3047 ATGTTAAAGAATGTCTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACA 3106
              || | ||||| || || |||||   | || ||| | |||||| | || |||||||| || |
Db        743 ATATCAAAGAGTGCCTATGGTTCATGCTTGAAGTTTTAATGACAAAGAATGAAAACAATA 802

Qy       3107 GTCACGCTTTTATCAGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGAC 3166
              | || || ||||| |  ||||||| ||| || || ||     ||| | ||||||||     |
Db        803 GCCATGCCTTTATGAAGAAGATGGCAGAGAACATCAAGTTAACCAGAGATGCCCAGTCTC 862

Qy       3167 CAGATGATGCAAAAATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCA 3226
              ||||||| | || |  ||||||||||||| ||| || |||||||| | |   | |  | | |
Db        863 CAGATGAATCCAAGACAAATGAAAAACTGTATACAGTATGTGATGTGGCTCTCTGTGTTA 922

Qy       3227 TCATGTCAAAGAGTACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTC 3286
              | |      || ||| ||   | ||   || |   ||| |||| ||     || | |
Db        923 TAAATAGTAAAAGTGCTTTGTGCA--ATGCAGATTCCCAAAGGACCCAGCCTCCAATGAA 980

Qy       3287 GTTTCTTCACTCAACCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAA 3346
              ||| | |||    |||||| ||| | ||| |||||    || | |||| |  |     || |
Db        981 ATTTTTACAC---ACCTGAAAAGGACTTCTGTAACGATAAGAGTTATATTTCAGAAGAGA 1037

Qy       3347 TGAAATCATTTTTCACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACA 3406
                | |  | | |        | ||||| || ||  |  |    || ||||| || || || | |
Db       1038 CAAGAGTACTTCTGTTAACAGGAAAGCCAAAGCCTGCTGGAGTACTAGGTGCAGTAAATA 1097

Qy       3407 AGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAA 3466
              ||||  ||| || |  ||| | |||||      |   |   |   || |||| ||
Db       1098 AGCCTTTATCAGCAACGGGAAGGAAACCCTATGTTAGAAGCACTGGCACTGAGACTGGAA 1157

Qy       3467 GCAATGCAAGCAGCAGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATA 3526
              |||||    |     |  ||| |  |  | | || |    |||        |||     | |
Db       1158 GCAATATTAATGTAAATTCAGAGCTGAACCCTTCAACCGGAAATCGATCAAGGGAACAGA 1217

Qy       3527 GTTCTGA 3533
              |||| ||
Db       1218 GTTCAGA 1224
```

RESULT 2
US-10-040-739-107
; Sequence 107, Application US/10040739
; Publication No. US20020173635A1
;   GENERAL INFORMATION:
;   APPLICANT: Jacobs, Kenneth
;   McCoy, John
;   LaVallie, Edward
;   Racie, Lisa
;   Merberg, David
;   Treacy, Maurice
;   Spaulding, Vikki
;   TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
;   NUMBER OF SEQUENCES: 1519
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Genetics Institute, Inc.
;   STREET: 87 CambridgePark Drive
;   CITY: Cambridge
;   STATE: Massachusetts
;   COUNTRY: U.S.A
;   ZIP: 02140
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy Disk
;   COMPUTER: IBM PC Compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/040,739
;   FILING DATE: 07-Jan-2002
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 09/036,520
;   FILING DATE: 03-JUN-1998
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Brown, Scott A.
;   REGISTRATION NUMBER: 32,724
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 498-8224
;   TELEFAX: (617) 876-5851
;   INFORMATION FOR SEQ ID NO: 107:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 439 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   SEQUENCE DESCRIPTION: SEQ ID NO: 107:
US-10-040-739-107

```
  Query Match             9.9%;  Score 411.4;  DB 13;  Length 439;
  Best Local Similarity   98.6%;  Pred. No. 1.5e-91;
  Matches  415;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       3216 CATGAATATCATCATGTCAAAGAGTACTACATACAGTTTGGAATCTCCTAAAGACCCGGT 3275
              ||||    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 CATGGCCTACATCATGTCAAAGAGTACTACATACAGTTTGGAATCTCCTAAAGACCCGGT 72

Qy       3276 ACTACCAGCTCGTTTCTTCACTCAACCTGACAAGAATTTCAGTAACACCAAAAATTATCT 3335
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 ACTACCAGCTCGTTTCTTCACTCAACCTGACAAGAATTTCAGTAACACCAAAAATTATCT 132

Qy       3336 GCCTCCTGAAATGAAATCATTTTTCACTCCTGGAAAACCTAAAACAACCAATGTTCTAGG 3395
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 GCCTCCTGAAATGAAATCATTTTTCACTCCTGGAAAACCTAAAACAACCAATGTTCTAGG 192

Qy       3396 AGCTGTTAACAAGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAAATCATCACGAAT 3455
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 AGCTGTTAACAAGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAAATCATCACGAAT 252

Qy       3456 GGAAACTGTAAGCAATGCAAGCAGCAGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGG 3515
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        253 GGAAACTGTAAGCAATGCAAGCAGCAGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGG 312
```

```
Qy       3516 GAGGCTTGATAGTTCTGAAATGGATCACAGTGAAAATGAAGATTACACAATGTCTTCACC 3575
Db        313 GAGGCTTGATAGTTCTGAAATGGATCACAGTGAAAATGAAGATTACACAATGTCTTCACC 372

Qy       3576 TTTGCCGGGGAAAAAAAGTGACAAGAGAGACGACTCTGATCTTGTAAGGTCTGAATTGGA 3635
Db        373 TTTGCCGGGGAAAAAAAGTGACAAGAGAGACGACTCTGATCTTGTAAGGTCTGAATCGGA 432

Qy       3636 G 3636
Db        433 G 433
```

RESULT 3
US-09-925-302-357
; Sequence 357, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 357
; LENGTH: 1217
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1141)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (1149)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (1157)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-357

```
  Query Match             6.1%;  Score 253.8;  DB 9;  Length 1217;
  Best Local Similarity  62.2%;  Pred. No. 3.4e-52;
  Matches  399;  Conservative    0;  Mismatches  242;  Indels    0;  Gaps    0;

Qy       2893 AATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGTGAAAAATTATTG 2952
Db          2 AATATCAGTATACGCAGGGAATACATTAAGCAGAATCCTATGGCTACTGAGAAATTATTA 61

Qy       2953 TCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCACATGACCCAGAT 3012
Db         62 TCACTGTTGCCTGAATATGTAGTTCCATACATGATTCACCTGCTAGCCCATGATCCAGAT 121

Qy       3013 TATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGTCTTTGGTTTGTT 3072
Db        122 TTTACAAGATCACAAGATGTTGATCAGCTTCGTGATATCAAAGAGTGCCTATGGTTCATG 181

Qy       3073 CTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATCAGAAAGATGGTA 3132
Db        182 CTTGAAGTTTTAATGACAAAGAATGAAAACAATAGCCATGCCTTTATGAAGAAGATGGCA 241

Qy       3133 GAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAAATGAATGAAAAA 3192
Db        242 GAGAACATCAAGTTAACCAGAGATGCCCAGTCTCCAGATGAATCCAAGACAAATGAAAAA 301

Qy       3193 CTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGTACTACATACAGT 3252
Db        302 CTGTATACAGTATGTGATGTGGCTCTCTGTGTTATAAATAGTAAAAGTGCTTTGTGCAAT 361

Qy       3253 TTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAACCTGACAAGAAT 3312
Db        362 GCAGATTCACCAAAGGACCCAGTCCTCCCAATGAAATTTTTACACAACCTGAAAAGGAC 421

Qy       3313 TTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTCACTCCTGGAAAA 3372
Db        422 TTCTGTAACGATAAGAGTTATATTTCAGAAGAGACAAGAGTACTTCTGTTAACAGGAAAG 481

Qy       3373 CCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCAGCAGGCAAGCAA 3432
Db        482 CCAAAGCCTGCTGGAGTACTAGGTGCAGTAAATAAGCCTTTATCAGCAACGGGAAGGAAA 541

Qy       3433 TCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGCAGCTCAAATCCA 3492
Db        542 CCCTATGTTAGAAGCACTGGCACTGAGACTGGAAGCAATATTAATGTAAATTCAGAGCTG 601

Qy       3493 AGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGA 3533
Db        602 AACCCTTCAACCGGAAATCGATCAAGGGAACAGAGTTCAGA 642
```

RESULT 4
US-09-918-995-3514/c
; Sequence 3514, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3514
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-3514

```
  Query Match             5.2%;  Score 217.6;  DB 11;  Length 471;
  Best Local Similarity  69.0%;  Pred. No. 1.8e-43;
  Matches  298;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;

Qy       1309 AAAGACAAATTGCTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAA 1368
Db        459 AAAGGCAAACTTCTGCATATTTATTATCAGAACAGCATTGACGACAAACTGTTGGTAGAG 400

Qy       1369 CGGATCTTTGCTCAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGC 1428
Db        399 AAAATCTTTGCTCAGTATCTTGTCCCCCACAACCTGGAAACAGAAGAGAGAATGAAATGC 340

Qy       1429 TTATATTACTTGTATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGG 1488
Db        339 TTATATTACTTATATGCTAGTTTGGATCCAAATGCTGTAAAAGCTCTCAACGAAATGTGG 280

Qy       1489 AAATGTCAAAATCTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCC 1548
Db        279 AAGTGTCAGAACATGCTTCGGAGCCATGTACGCGAACTATTGGATTTGCACAAGCAGCCT 220

Qy       1549 AAAACAGATGCCAGTGTCAAGGCCATATTTCAAAAGTGATGGTTATTACAAGAAATTTA 1608
Db        219 ACATCAGAGGCTAACTGTTCTGCCATGTTTGGAAAACTGATGACCATAGCAAAGAATTTG 160

Qy       1609 CCTGATCCTGGTAAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGAT 1668
Db        159 CCTGACCCCGGGAAAGCACAAGATTTGTGAAGAAATTTAACCAGGTTCTCGGCGATGAT 100

Qy       1669 GAGAAAATAAGAAAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCT 1728
```

```
Db          99 GAGAAACTTCGGTCTCAGTTGGAGTTATTAATTAGCCCAACCTGTTCTTGCAAACAAGCA 40

Qy        1729 GAAGGTTGTGTG 1740
               ||   |||||||
Db          39 GATATTTGTGTG 28

RESULT 5
US-09-879-536-682/c
; Sequence 682, Application US/09879536
; Patent No. US20020144298A1
; GENERAL INFORMATION:
;  APPLICANT: Endege, Wilson O.
;  APPLICANT:  Steinmann, Kathleen E.
;  APPLICANT:  Astle, Jon H.
;  APPLICANT:  Burgess, Christopher C.
;  APPLICANT:  Bushnell, Steven E.
;  APPLICANT:  Carroll III, Eddie
;  APPLICANT:  Catino, Theodore J.
;  APPLICANT:  Derti, Adnan
;  APPLICANT:  Ford, Donna M.
;  APPLICANT:  Lewis, Marcia E.
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Schlegel, Robert
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION:  PRODUCTS
;  FILE REFERENCE: CCD-257 (US)
;  CURRENT APPLICATION NUMBER: US/09/879,536
;  CURRENT FILING DATE:  2001-09-21
;  PRIOR APPLICATION NUMBER: US 60/088,801
;  PRIOR FILING DATE: 1998-06-10
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 682
;   LENGTH: 530
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-879-536-682

  Query Match             4.9%;  Score 203.6;  DB 10;  Length 530;
  Best Local Similarity   98.1%;  Pred. No. 5.7e-40;
  Matches  206;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         104 TAAAGATGGTTGTGAAAACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGC 163
               || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         218 TATAGATGGTTGTGAAAACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGC 159

Qy         164 TTTATTTAAACCTAGCTTTACATCTTGCTTCAGATTTTTTCTCAAGCATCCTGGTAAAG 223
               ||||||||||||||||||||||||||||||||||||||| ||||||||||||| ||| |
Db         158 TTTATTTAAACCTAGCTTTACATCTTGCTTCAGATTTTTCCTCAAGCATCCTGATAAGG 99

Qy         224 ATGTTCGCTTACTGGTAGCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAG 283
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          98 ATGTTCGCTTACTGGTAGCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAG 39

Qy         284 CTCCTTACACATCCCCTGATAAACTAAAGG 313
               ||||||||||||||||||||||||||||||
Db          38 CTCCTTACACATCCCCTGATAAACTAAAGG 9

RESULT 6
US-09-918-995-33817
; Sequence 33817, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33817
;   LENGTH: 433
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-918-995-33817

  Query Match             3.9%;  Score 164;  DB 11;  Length 433;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-30;
  Matches  164;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        4009 ACGGAACAGAAGTCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCA 4068
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           3 ACGGAACAGAAGTCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCA 62

Qy        4069 GAATCTCCTGAATCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGA 4128
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          63 GAATCTCCTGAATCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGA 122

Qy        4129 AGACCATCAAAAACGCCATCACCATCACAACCAAAAAAAAATGT 4172
               ||||||||||||||||||||||||||||||||||||||||||||
Db         123 AGACCATCAAAAACGCCATCACCATCACAACCAAAAAAAAATGT 166

RESULT 7
US-09-918-995-20720
; Sequence 20720, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 20720
;   LENGTH: 413
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-918-995-20720

  Query Match             2.3%;  Score 96.4;  DB 11;  Length 413;
  Best Local Similarity   74.7%;  Pred. No. 1.8e-13;
  Matches  121;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy         107 AGATGGTTGTGAAAACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTT 166
               ||||||| |||||||| ||||||||||||||| |||||||| ||||| |||||  |||  |
Db         252 AGATGGTAGTGAAAACCTTTATGGATATGGATCAGGACTCAGAAGATGAAAAACAGCAGT 311

Qy         167 ATTTAAACCTAGCTTTACATCTTGCTTCAGATTTTTTCTCAAGCATCCTGGTAAAGATG 226
               || |     ||||| || |||||||| || || || || |||| | | ||     |||||||
Db         312 ATCTCCCACTAGCCTTGCATCTTGCATCTGAATTCTTCCTCAGGAACCCCAATAAAGATG 371

Qy         227 TTCGCTTACTGGTAGCCTGCTGCCTTGCTGATATTTTCAGGA 268
               | ||   | || ||||| || ||   | |||||||| ||  | |
Db         372 TGCGTCTCCTTGTAGCATGTTGTTTGGCTGATATCTTTCGTA 413

RESULT 8
US-09-728-445-172
; Sequence 172, Application US/09728445
; Patent No. US20020102543A1
; GENERAL INFORMATION:
;  APPLICANT: Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020102543A1el Mutated Mammalian Cells and
;  TITLE OF INVENTION:  Animals
```

```
;  FILE REFERENCE: LEX-0102-USA
;  CURRENT APPLICATION NUMBER: US/09/728,445
;  CURRENT FILING DATE:  2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/168,358
;  PRIOR FILING DATE: 1999-12-01
;  NUMBER OF SEQ ID NOS: 891
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 172
;   LENGTH: 331
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(331)
;   OTHER INFORMATION: n = A,T,C or G
US-09-728-445-172

  Query Match             2.1%;  Score 87.4;  DB 10;  Length 331;
  Best Local Similarity   85.1%;  Pred. No. 2.7e-11;
  Matches   97;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy        190 GCTTCAGATTTTTTCTCAAGCATCCTGGTAAAGATGTTCGCTTACTGGTAGCCTGCTGC 249
              | ||| || |   || |||| |||||||| |||||||||||| |||||||| || ||||||
Db        216 GATTCTGACTCCTTCCTCANGCATCCTGATAAAGATGTTCGTTTACTGGTGGCTTGCTGC 275

Qy        250 CTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCTGAT 303
              |||||||||||||||||||||||||||||| |||||||||| ||     |||
Db        276 CTTGCTGATATTTTCAGGATTTATGCTCCTGAGGCTCCTTACACGTCGTGAGAT 329


RESULT 9
US-10-027-632-105871
; Sequence 105871, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 105871
;   LENGTH: 642
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-105871

  Query Match             1.8%;  Score 76.8;  DB 13;  Length 642;
  Best Local Similarity   90.0%;  Pred. No. 1.8e-08;
  Matches   81;  Conservative    1;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       2388 TCTTCTCATGAATGATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGA 2447
              | |||||:|  ||  |  |||||||||||||||||||||||||||||||||||||||||
Db        372 TTTTCTCRTTTATTTTAGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGA 431

Qy       2448 AGAAGTATCTCCTGAGACAATGGTCAAAAT 2477
              |||||||||||||||||||||||||||| |
Db        432 AGAAGTATCTCCTGAGACAATGGTCAAAGT 461


RESULT 10
US-10-027-632-19016/c
; Sequence 19016, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19016
;   LENGTH: 739
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-19016

  Query Match             1.8%;  Score 76.8;  DB 13;  Length 739;
  Best Local Similarity   90.0%;  Pred. No. 1.9e-08;
  Matches   81;  Conservative    1;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       2388 TCTTCTCATGAATGATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGA 2447
              | |||||:|  ||  |  |||||||||||||||||||||||||||||||||||||||||
Db        272 TTTTCTCRTTTATTTTAGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGA 213

Qy       2448 AGAAGTATCTCCTGAGACAATGGTCAAAAT 2477
              |||||||||||||||||||||||||||| |
Db        212 AGAAGTATCTCCTGAGACAATGGTCAAAGT 183


RESULT 11
US-10-311-455-1931/c
; Sequence 1931, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Der
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1931
;   LENGTH: 14006
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

```
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 8289, 8310, 8313
;   OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1931

  Query Match             1.5%;  Score 64.4;  DB 12;  Length 14006;
  Best Local Similarity   48.4%;  Pred. No. 0.00015;
  Matches  179;  Conservative    0;  Mismatches  191;  Indels    0;  Gaps    0;

Qy       3800 AAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCAAAAA 3859
Db       2484 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2425

Qy       3860 AGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAGAAGAGCCAA 3919
Db       2424 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACCAAATAAAAAAAAAA 2365

Qy       3920 CAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGG 3979
Db       2364 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATATAAAAAAAAAAAAAAAAAAAAA 2305

Qy       3980 AGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAACAGCACC 4039
Db       2304 AACGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAACG 2245

Qy       4040 GAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAATCTAGTGCAATTGAATCCA 4099
Db       2244 AAAAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAA 2185

Qy       4100 CACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAAACGCCATCACCATCACAAC 4159
Db       2184 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATCAAAAAAAAAAAAAAAAAAAAAAA 2125

Qy       4160 CAAAAAAAAA 4169
Db       2124 AAAAAAAAAA 2115

RESULT 12
US-10-311-455-1670/c
; Sequence 1670, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1670
;   LENGTH: 6668
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 1936
;   OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1670

  Query Match             1.5%;  Score 62.8;  DB 12;  Length 6668;
  Best Local Similarity   48.1%;  Pred. No. 0.00023;
  Matches  178;  Conservative    0;  Mismatches  192;  Indels    0;  Gaps    0;

Qy       3800 AAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCAAAAA 3859
Db       3339 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAAAACAAAAACAAAAAAAAAA 3280

Qy       3860 AGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAGAAGAGCCAA 3919
Db       3279 AAAAAAAAACAAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAAA 3220

Qy       3920 CAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGG 3979
Db       3219 AAAAAAAAAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAAAAC 3160

Qy       3980 AGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAACAGCACC 4039
Db       3159 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAA 3100

Qy       4040 GAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAATCTAGTGCAATTGAATCCA 4099
Db       3099 CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAAAAAAAAA 3040

Qy       4100 CACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAAACGCCATCACCATCACAAC 4159
Db       3039 AAAAAAAAAAAAAAAAAAACAAAAACAAAAAAAAAAAAAAAAAAAAAAACAAAAAAAAA 2980

Qy       4160 CAAAAAAAAA 4169
Db       2979 AAAAAAAAAA 2970

RESULT 13
US-10-001-843-45
; Sequence 45, Application US/10001843
; Publication No. US20020132255A1
; GENERAL INFORMATION:
;  APPLICANT: Salceda, Susana
;  APPLICANT:  Macina, Roberto
;  APPLICANT:  Recipon, Herve
;  APPLICANT:  Cafferkey, Robert
;  APPLICANT:  Sun, Yongming
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Turner, Leah
;  TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and
;  FILE REFERENCE: DEX-0267
;  CURRENT APPLICATION NUMBER: US/10/001,843
;  CURRENT FILING DATE:  2001-11-20
;  PRIOR APPLICATION NUMBER: 60/249,992
;  PRIOR FILING DATE: 2000-11-20
;  NUMBER OF SEQ ID NOS: 218
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45
;   LENGTH: 1267
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (358)..(358)
;   OTHER INFORMATION: a, c, g or t
;   NAME/KEY: misc_feature
;   LOCATION: (478)..(478)
;   OTHER INFORMATION: a, c, g or t
US-10-001-843-45

  Query Match             1.5%;  Score 61.2;  DB 13;  Length 1267;
  Best Local Similarity   45.9%;  Pred. No. 0.0002;
  Matches  210;  Conservative    0;  Mismatches  248;  Indels    0;  Gaps    0;

Qy       3714 ACAGGAACAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGA 3773
Db        744 AAAAAAATAAAAGAGGAAAAACAAACAGAAAAGACAAAAAAACAAAAAAGAACAAAAAAA 803

Qy       3774 ATCTGATGAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGA 3833
```

```
Db        804 AACAGACAAAGAAAAAAAAAAAAAGAAAAAAAAAAACAAAAAAGAAAAAAAACAGGAACA 863

Qy       3834 AGATGAACAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGG 3893

Db        864 ATAAAAAAAAAGAAAAACACAAAACAACAGAACAACAGAAGAAAAAAAAAAGAAGAGAGA 923

Qy       3894 TGGAGGTACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAA 3953

Db        924 GAGAAAAAAACAAAAAGAAAAAAAAAAAAAACAAAAAAGAACAAAAAAAAAGAAAAAAAA 983

Qy       3954 ATCTGGACCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGA 4013

Db        984 CAAAAAACAAAAAACAAGAAAAAAAAAGAAAAAAAAACAAAAAAGCAAAAAACAAAGAAA 1043

Qy       4014 ACAGAAGTCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATC 4073

Db       1044 GAGAAGAGGAAAAAATAAAGAGCAAAAAAACAAAAAAAAAAAGAAAAATGACAAAAAACA 1103

Qy       4074 TCCTGAATCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACC 4133

Db       1104 CGAAAAAACAAGATACAACAAAACAAGAAAAAGAAACAAAAAGAAAAAAGAAGAAACACA 1163

Qy       4134 ATCAAAAACGCCATCACCATCACAACCAAAAAAAAATG 4171

Db       1164 AAGAAAAACAAAAAAAAACAGAGAAGAAAGAAAAAAAG 1201
```

RESULT 14
US-10-240-453-54/c
; Sequence 54, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
; TITLE OF INVENTION: With DNA TRanscription
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240,453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 54
; LENGTH: 9539
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-54

```
  Query Match             1.5%;  Score 60.6;  DB 12;  Length 9539;
  Best Local Similarity   43.7%;  Pred. No. 0.001;
  Matches  468;  Conservative    0;  Mismatches  589;  Indels   14;  Gaps    4;

Qy       3077 AAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATCAGAAAGATGGTAGAAA 3136

Db       1099 AAAAAAAAAACGAAAAAACGAAAAAAATAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAT 1040

Qy       3137 ATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAAATGAATGAAAAACTGT 3196

Db       1039 AAAAAAAAAAAAAAAAAAATAAACAAAAAAAAAAAAAAAAAAAATAAAAAAATTAAAAAAA 980

Qy       3197 ACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGTACTACATACAGTTTGG 3256

Db        979 AAAAAAAAATAAATAATAAAAAAAAAAAAATAAAAAATAAATATAAACAAAAAAAAAAAA 920

Qy       3257 AATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAACCTGA--CAAGAATTT 3314

Db        919 CCGAAAATAATAACTCACACCTATAATCCCAATACTTTAAAAAACTAAAATAAAAAAATC 860

Qy       3315 CAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTCACTCCTGGAAAACC 3374

Db        859 ACTTAAAACCAAAAATTCAAAACCAACCTAAACAACAAAATCATACCCCATTTCTACAAA 800

Qy       3375 TAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCAT--CAGCAGGCAAGCAA 3432

Db        799 AAAAAAACAAAAACTAACCAAACCTAATAACACACACCTATAATCCCAACTACTCAAAAA 740

Qy       3433 TCTCAGACCAAATCATCACGAATGGAAA--CTGTAAGCAATGCAAGCAGCAGCTCAAATC 3490

Db        739 ACTAAAACAAAAAAATCACTACCGAAAAACAATAATTACAATAAACCGAAATCGTACCAC 680

Qy       3491 CAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGATCACAGTGAAA 3550

Db        679 TACACTCCAACCTAACAACAAAAAAAACCCTACCTCAAAAAAAAATATATATAAATAAAA 620

Qy       3551 ATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAGAGAGACGACT 3610

Db        619 AAAAAAAAAATAAAATATATAAAAAAA--------ATAAAAAAAAAAAATAAAACAAAA 568

Qy       3611 CTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACGCCCGTCACAG 3670

Db        567 AAAAAAAAAAAACAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAAATCAAAAAAAA 508

Qy       3671 AACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAACAGAAACCTA 3730

Db        507 AAAAAAAAAATAAAAAAAAAAAAAAAAACAAAAAAAAATAAAAAAAAAATAAAAAAAAA 448

Qy       3731 AAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGT 3790

Db        447 AAAAATAAAAAAAAAAAAAACAAAAAAAAAATAAAAAAAAAAAAAACGAAAAAAAAACAA 388

Qy       3791 GGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTC 3850

Db        387 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAATAAAAC 328

Qy       3851 CGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAG 3910

Db        327 AAATAAAAATAAAAAAAAAACGAAAAAAAAAAAAAAAAAAAAAAACAAACGACAAAAAAAA 268

Qy       3911 AAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCAC 3970

Db        267 AAAAACGAAAAAAAAAATAAAAAAAAAACACACAAAAATAAAACGAAAAAATACAAAAA 208

Qy       3971 CAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCA 4030

Db        207 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTAAAATAAAA 148

Qy       4031 AACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAATCTAGTGCAA 4090

Db        147 AAAAAAAAAAAATAAAAAAATAAAAAAAAAAAATATAAAAACTAAAATAATAAAATAAAA 88

Qy       4091 TTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAAA 4141

Db         87 AAAAAAAAAAAAAAAAAAAAAAAACAAATAAAAAAAAATAAAAAAAAAA 37
```

RESULT 15
US-10-239-676-52/c
; Sequence 52, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian

```
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 52
;   LENGTH: 9539
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-52

  Query Match             1.5%;  Score 60.6;  DB 14;  Length 9539;
  Best Local Similarity   43.7%;  Pred. No. 0.001;
  Matches  468;  Conservative    0;  Mismatches  589;  Indels   14;  Gaps    4;

Qy       3077 AAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATCAGAAAGATGGTAGAAA 3136
Db       1099 AAAAAAAAAACGAAAAAACGAAAAAAATAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAT 1040

Qy       3137 ATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAAATGAATGAAAAACTGT 3196
Db       1039 AAAAAAAAAAAAAAAAAAATAAACAAAAAAAAAAAAAAAAAAATAAAAAAATTAAAAAAA 980

Qy       3197 ACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGTACTACATACAGTTTGG 3256
Db        979 AAAAAAAAATAAATAATAAAAAAAAAAAAATAAAAAATAAATATAAACAAAAAAAAAAAA 920

Qy       3257 AATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAACCTGA--CAAGAATTT 3314
Db        919 CCGAAAATAATAACTCACACCTATAATCCCAATACTTTAAAAAACTAAAATAAAAAAATC 860

Qy       3315 CAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTCACTCCTGGAAAACC 3374
Db        859 ACTTAAAACCAAAAATTCAAAACCAACCTAAACAACAAAATCATACCCCATTTCTACAAA 800

Qy       3375 TAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCAT--CAGCAGGCAAGCAA 3432
Db        799 AAAAAAACAAAAACTAACCAAACCTAATAACACACACCTATAATCCCAACTACTCAAAAA 740

Qy       3433 TCTCAGACCAAATCATCACGAATGGAAA--CTGTAAGCAATGCAAGCAGCAGCTCAAATC 3490
Db        739 ACTAAAACAAAAAAATCACTACCGAAAAACAATAATTACAATAAACCGAAATCGTACCAC 680

Qy       3491 CAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGATCACAGTGAAA 3550
Db        679 TACACTCCAACCTAACAACAAAAAAAACCCTACCTCAAAAAAAAATATATATAAATAAAA 620

Qy       3551 ATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAGAGAGACGACT 3610
Db        619 AAAAAAAAAATAAAATATATAAAAAAA--------ATAAAAAAAAAAAATAAAAACAAAA 568

Qy       3611 CTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACGCCCGTCACAG 3670
Db        567 AAAAAAAAAAAACAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAATCAAAAAAAAA 508

Qy       3671 AACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAACAGAAACCTA 3730
Db        507 AAAAAAAAAATAAAAAAAAAAAAAAAAACAAAAAAAAATAAAAAAAAAATAAAAAAAAAA 448

Qy       3731 AAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGT 3790
Db        447 AAAAATAAAAAAAAAAAAAACAAAAAAAAAATAAAAAAAAAAAAAACGAAAAAAAACAA 388

Qy       3791 GGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTC 3850
Db        387 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAATAAAAAC 328

Qy       3851 CGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAG 3910
Db        327 AAATAAAAATAAAAAAAAAACGAAAAAAAAAAAAAAAAAAAAAACAAACGACAAAAAAAA 268

Qy       3911 AAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCAC 3970
Db        267 AAAAACGAAAAAAAAAATAAAAAAAAAACACACAAAAATAAAACGAAAAAATACAAAAA 208

Qy       3971 CAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCA 4030
Db        207 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTAAAATAAAA 148

Qy       4031 AACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAATCTAGTGCAA 4090
Db        147 AAAAAAAAAAAATAAAAAAATAAAAAAAAAAAATATAAAAACTAAAATAATAAAATAAAA 88

Qy       4091 TTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAAA 4141
Db         87 AAAAAAAAAAAAAAAAAAAAAAAAACAAATAAAAAAAAATAAAAAAAAAAA 37

Search completed: September 24, 2003, 20:16:04
Job time : 919 secs
```

wpa

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

```
Run on:         September 24, 2003, 09:42:57 ; Search time 223 Seconds
                                          (without alignments)
                                          8259.605 Million cell updates/sec

Title:          US-09-512-581B-3
Perfect score:  4173
Sequence:       1 atggctcattcaaagactag..........cacaaccaaaaaaaaatgtg 4173

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters:      1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                2:  /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                6:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	203.6	4.9	530	3	US-09-328-111-682	Sequence 682, App
c	2	99.8	2.4	7218	1	US-08-232-463-14	Sequence 14, Appl
	3	54.6	1.3	6755	3	US-08-931-999-4	Sequence 4, Appli
c	4	51.4	1.2	72604	4	US-09-268-992-7	Sequence 7, Appli
c	5	51.4	1.2	72604	4	US-09-657-474-7	Sequence 7, Appli
	6	44.8	1.1	5433	3	US-08-929-329-1	Sequence 1, Appli
	7	44	1.1	289	3	US-09-007-005-17	Sequence 17, Appl
	8	44	1.1	289	3	US-09-244-796-17	Sequence 17, Appl
c	9	43.6	1.0	435	4	US-09-134-001C-1924	Sequence 1924, Ap
	10	43.6	1.0	1089	4	US-09-134-001C-1833	Sequence 1833, Ap
	11	43.6	1.0	5552	3	US-08-155-888-1	Sequence 1, Appli
	12	42.4	1.0	2169	4	US-09-434-408-3	Sequence 3, Appli
	13	42.4	1.0	4766	5	PCT-US93-07261-10	Sequence 10, Appl
	14	42	1.0	1298	3	US-08-948-705-3	Sequence 3, Appli
	15	42	1.0	1298	4	US-09-510-543-3	Sequence 3, Appli
	16	42	1.0	2447	2	US-09-014-969-14	Sequence 14, Appl
	17	41.8	1.0	240	1	US-08-628-417-6	Sequence 6, Appli
	18	41.8	1.0	1236	2	US-08-741-134-5	Sequence 5, Appli
	19	41.6	1.0	2859	2	US-08-506-340A-2	Sequence 2, Appli
c	20	41.4	1.0	14066	4	US-09-601-198-56	Sequence 56, Appl
c	21	41.2	1.0	339	4	US-09-134-001C-1851	Sequence 1851, Ap
	22	41.2	1.0	543	6	5273901-6	Patent No. 5273901
	23	40.8	1.0	454	2	US-08-623-906A-6	Sequence 6, Appli
	24	40.8	1.0	9139	4	US-09-322-478-22	Sequence 22, Appl
	25	40.4	1.0	376	2	US-08-623-906A-18	Sequence 18, Appl
	26	40.4	1.0	949	4	US-09-489-847-35	Sequence 35, Appl
	27	40.4	1.0	1798	4	US-09-797-906-1	Sequence 1, Appli
	28	40.2	1.0	696	3	US-09-461-697-193	Sequence 193, App
	29	40.2	1.0	699	3	US-09-461-697-191	Sequence 191, App
	30	40.2	1.0	717	3	US-09-461-697-189	Sequence 189, App
	31	40.2	1.0	774	3	US-09-461-697-187	Sequence 187, App
	32	40.2	1.0	819	3	US-09-461-697-185	Sequence 185, App
	33	40.2	1.0	1669	3	US-09-461-697-184	Sequence 184, App
	34	40	1.0	277	3	US-09-007-005-3	Sequence 3, Appli
	35	40	1.0	277	3	US-09-244-796-3	Sequence 3, Appli
	36	40	1.0	1696	4	US-09-835-811-1	Sequence 1, Appli
	37	40	1.0	2025	4	US-09-107-532A-2547	Sequence 2547, Ap
c	38	40	1.0	6693	4	US-08-961-527-195	Sequence 195, App
	39	39.6	0.9	1275	4	US-09-286-981B-28	Sequence 28, Appl
	40	39.4	0.9	2672	3	US-09-214-564A-5	Sequence 5, Appli
	41	39.4	0.9	2815	3	US-09-214-564A-1	Sequence 1, Appli
	42	39.4	0.9	3275	4	US-09-370-838-151	Sequence 151, App
c	43	39.4	0.9	4185	4	US-09-417-485D-7	Sequence 7, Appli
c	44	39.4	0.9	10640	4	US-09-417-485D-5	Sequence 5, Appli
c	45	39.4	0.9	1664976	4	US-08-916-421B-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-09-328-111-682/c
; Sequence 682, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
;  APPLICANT: Endege, Wilson O.
;  APPLICANT: Steinmann, Kathleen E.
;  APPLICANT: Astle, Jon H.
;  APPLICANT: Burgess, Christopher C.
;  APPLICANT: Bushnell, Steven E.
;  APPLICANT: Carroll III, Eddie
;  APPLICANT: Catino, Theodore J.
;  APPLICANT: Derti, Adnan
;  APPLICANT: Ford, Donna M.
;  APPLICANT: Lewis, Marcia E.
;  APPLICANT: Monahan, John E.
;  APPLICANT: Schlegel, Robert
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS
;  FILE REFERENCE: CCD-257 (US)
;  CURRENT APPLICATION NUMBER: US/09/328,111
;  CURRENT FILING DATE: 1999-06-08
;  EARLIER APPLICATION NUMBER: US 60/088,801
;  EARLIER FILING DATE: 1998-06-10
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 682
;   LENGTH: 530
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-328-111-682

  Query Match             4.9%;  Score 203.6;  DB 3;  Length 530;
  Best Local Similarity   98.1%;  Pred. No. 1.2e-44;
  Matches  206;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        104 TAAAGATGGTTGTGAAAACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGC 163
              || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        218 TATAGATGGTTGTGAAAACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGC 159

Qy        164 TTTATTTAAACCTAGCTTTACATCTTGCTTCAGATTTTTTCTCAAGCATCCTGGTAAAG 223
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||| ||| |
Db        158 TTTATTTAAACCTAGCTTTACATCTTGCTTCAGATTTTTCCTCAAGCATCCTGATAAGG 99

Qy        224 ATGTTCGCTTACTGGTAGCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAG 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         98 ATGTTCGCTTACTGGTAGCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAG 39

Qy        284 CTCCTTACACATCCCCTGATAAACTAAAGG 313
              ||||||||||||||||||||||||||||||
```

```
Db         38 CTCCTTACACATCCCCTGATAAACTAAAGG 9

RESULT 2
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             2.4%;  Score 99.8;  DB 1;  Length 7218;
  Best Local Similarity   4.3%;  Pred. No. 2.6e-16;
  Matches   17;  Conservative  256;  Mismatches  118;  Indels    0;  Gaps    0;

Qy       3700 TTGACTAAGTTGGTACAGGAACAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGC 3759
               | ||  || ||||||::::: :::::   ::::: ::  :: ::::  ::::::::::
Db       1450 TAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1391

Qy       3760 CATACGGCTTCAGAATCTGATGAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGAT 3819
                : : ::    ::::  :: ::: :: :: ::   ::::::::::::   ::::::::
Db       1390 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1331

Qy       3820 ATATTAGAAAATGAAGATGAACAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCA 3879
              : :  :::::: ::::: ::: :::: ::   :  ::::::::: ::::::::  ::  :
Db       1330 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1271

Qy       3880 CCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAA 3939
                 ::::      :: ::::: : :  :::::::::  :: :: :::::       ::::::
Db       1270 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1211

Qy       3940 GGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAA 3999
              ::::: :::::::::   :::     :: :  ::::::::::::::::::::::: ::
Db       1210 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1151

Qy       4000 AGTGGAAATACGGAACAGAAGTCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAG 4059
              :: ::::: : :::: :::::    ::::: ::: :: :   ::: :   :::::::: : ::
Db       1150 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1091

Qy       4060 CAGAGAGCAGAATCTCCTGAATCTAGTGCAA 4090
                :::::: ::::        :::   |  ||||
Db       1090 RRRRRRRRRRRRRRRRRRRRRRRRATCGCAA 1060

RESULT 3
US-08-931-999-4
; Sequence 4, Application US/08931999
; Patent No. 6043219
;  GENERAL INFORMATION:
;    APPLICANT:  Iandolo, John J.
;    APPLICANT:  Crupper, Scott S.
;    TITLE OF INVENTION:  Broad Spectrum Chemotherapeutic Peptide
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hovey, Williams, Timmons & Collins
;      STREET:  2405 Grand Boulevard, Suite 400
;      CITY:  Kansas City
;      STATE:  Missouri
;      COUNTRY:  U.S.A.
;      ZIP:  64108
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/931,999
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/710,561
;      FILING DATE:  19-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Collins, John M.
;      REGISTRATION NUMBER:  26,262
;      REFERENCE/DOCKET NUMBER:  25043-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  816/474-9050
;      TELEFAX:  816/474-9057
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6755 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Staphylococcus aureus
;      STRAIN:  UT0007
US-08-931-999-4

  Query Match             1.3%;  Score 54.6;  DB 3;  Length 6755;
  Best Local Similarity   46.6%;  Pred. No. 0.00029;
  Matches  174;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy       3797 AGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCAA 3856
              | |||||  | |     ||  | |        ||| |    ||     || | || |  |      |
Db       6204 ACGAAAAACGACAAGAAACAAAACACAAAGAGAGAAAAAGAAAAGAAAAAAAACGCAGGA 6263

Qy       3857 AAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAGC 3916
               |||     |  || |     |         |||| | |      |   |    |||||||  ||
Db       6264 GAAACAAAAGCAGGGAAACAAGCAAAAAAACCCGCAAAGACGAGCCCACAAAAGAGGAAG 6323
```

```
Qy       3917 CAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAGG 3976
               |||   |||||      ||||||  |||| ||||| ||    ||| |  ||      | ||
Db       6324 AAACCGCGAAAAAAAGAAAAAAAAAAAGCCAAAAACAAAAAGGAACAACAAGCAAAAAGA 6383

Qy       3977 AGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAACAGC 4036
               || || ||     ||||||| |  | |     | | |  ||||| ||| | ||||    |||
Db       6384 AGAAGAAAAGGGAAGAAAAAAGAGGAAAACAGAGAGAACAAAAGCCAAAAAAACAACGAA 6443

Qy       4037 ACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAATCTAGTGCAATTGAAT 4096
                | |||     |||    |   |     | |      | |        || |   |  |||   ||
Db       6444 GCGGAGAAGAAAGAGCAAAAACAAACACGCAAAAGGGCAAAAAAGCACGACCAAACAAAA 6503

Qy       4097 CCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAAACGCCATCACCATCAC 4156
                     |       |||| || ||  |   |   |   |        |  ||| |    ||  |   |   |
Db       6504 GACAAACAAGCACAACAAAGCGAACAACACAACCCAAAAAAAAGAGAACAGAAAAACAAA 6563

Qy       4157 AACCAAAAAAAAA 4169
               || |  || || |
Db       6564 CACGAGAACAACA 6576
```

RESULT 4
US-09-268-992-7/c
; Sequence 7, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 72604
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: all n positions
;   OTHER INFORMATION: n=a, c, g, or t
US-09-268-992-7

```
  Query Match             1.2%;  Score 51.4;  DB 4;  Length 72604;
  Best Local Similarity   46.1%;  Pred. No. 0.0076;
  Matches  172;  Conservative    0;  Mismatches  201;  Indels    0;  Gaps    0;

Qy       3693 GGATGACTTGACTAAGTTGGTACAGGAACAGAAACCTAAAGGCAGTCAGCGAAGTCGGAA 3752
               |||| ||||||        ||| |  |   ||        ||   ||| ||      |    |||
Db       7381 GGATAACTTGAGCCCAGGGGTTCTAGGCTAGCCTGAGCAACATAGTGAGAACTGGTCGAA 7322

Qy       3753 AAGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGTGGCCTGAGGAAAAGAGGCTCAA 3812
               |  | |     |  |    | |    ||   | || |          || || | ||     ||
Db       7321 AGAAAGAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 7262

Qy       3813 AGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCAAAAAAGGGTAAAAGAGG 3872
               || |   |      || ||    | || |  |   |||              ||   | ||  |||||
Db       7261 AGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAGAAAGAAA 7202

Qy       3873 CCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAGCCAACAATGAAAACTTC 3932
                  |   |   ||          |      |  | |  || | | | ||   ||  | | | ||
Db       7201 GAAAGGAAGGAAGGAAGGAAGGAAGGAGAAAGAAAGAAAGGAAGGAAGGAAGGAAGGAAG 7142
```

```
Qy       3933 TAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGGAGGAAGAAGAAGA 3992
               || ||||||| ||| |||||     ||      ||      |||       || ||||  ||||
Db       7141 GAAGGAAGGAAGGAAAGAAAAAGAAAGAAAGAAAGAGAAAGAAAGAAAGAAAGAGAAAGA 7082

Qy       3993 AAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAACAGCACCGAGTGTCAAGGAG 4052
               |||| |     ||||||| |||    |||   | ||||  |    |    |   ||| |   || ||
Db       7081 AAGAAAGGAAGGAAATAGAGAAAGAAAGAACGAAAGAGAGAAAGAGAGAGAGAGAAAGAA 7022

Qy       4053 AGCACAGCAGAGA 4065
               || |    |||||
Db       7021 AGAAAGAGAGAGA 7009
```

RESULT 5
US-09-657-474-7/c
; Sequence 7, Application US/09657474
; Patent No. 6399762
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/657,474
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/268,992
; PRIOR FILING DATE: 1999-03-16
; PRIOR APPLICATION NUMBER: 09/236,134
; PRIOR FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/106,056
; PRIOR FILING DATE: 1998-10-28
; PRIOR APPLICATION NUMBER: 60/088,312
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/078,044
; PRIOR FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 72604
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: all n positions
;   OTHER INFORMATION: n=a, c, g, or t
US-09-657-474-7

```
  Query Match             1.2%;  Score 51.4;  DB 4;  Length 72604;
  Best Local Similarity   46.1%;  Pred. No. 0.0076;
  Matches  172;  Conservative    0;  Mismatches  201;  Indels    0;  Gaps    0;

Qy       3693 GGATGACTTGACTAAGTTGGTACAGGAACAGAAACCTAAAGGCAGTCAGCGAAGTCGGAA 3752
               |||| ||||||        ||| |  |   ||        ||   ||| ||      |    |||
Db       7381 GGATAACTTGAGCCCAGGGGTTCTAGGCTAGCCTGAGCAACATAGTGAGAACTGGTCGAA 7322

Qy       3753 AAGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGTGGCCTGAGGAAAAGAGGCTCAA 3812
               |  | |     |  |    | |    ||   | || |          || || | ||     ||
Db       7321 AGAAAGAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 7262

Qy       3813 AGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCAAAAAAGGGTAAAAGAGG 3872
               || |   |      || ||    | || |  |   |||              ||   | ||  |||||
Db       7261 AGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAGAAAGAAA 7202

Qy       3873 CCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAGCCAACAATGAAAACTTC 3932
                  |   |   ||          |      |  | |  || | | | ||   ||  | | | ||
Db       7201 GAAAGGAAGGAAGGAAGGAAGGAAGGAGAAAGAAAGAAAGGAAGGAAGGAAGGAAGGAAG 7142

Qy       3933 TAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGGAGGAAGAAGAAGA 3992
               || ||||||| ||| |||||     ||      ||      |||       || ||||  ||||
Db       7141 GAAGGAAGGAAGGAAAGAAAAAGAAAGAAAGAAAGAGAAAGAAAGAAAGAAAGAGAAAGA 7082

Qy       3993 AAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAACAGCACCGAGTGTCAAGGAG 4052
               |||| |     ||||||| |||    |||   | ||||  |    |    |   ||| |   || ||
```

```
Db       7081 AAGAAAGGAAGGAAATAGAGAAAGAAAGAACGAAAGAGAGAAAGAGAGAGAGAGAAAGAA 7022

Qy       4053 AGCACAGCAGAGA 4065
              || |    |||||
Db       7021 AGAAAGAGAGAGA 7009

RESULT 6
US-08-929-329-1
; Sequence 1, Application US/08929329
; Patent No. 6120770
;  GENERAL INFORMATION:
;    APPLICANT:  Adams, John H
;    APPLICANT:  Dalton, John P
;    APPLICANT:  Kappe, Stefan
;    TITLE OF INVENTION:  Plasmodium Proteins Useful for Preparing
;    TITLE OF INVENTION:  Vaccine Compositions
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Barnes & Thornburg
;      STREET:  11 S Meridian
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46204
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/929,329
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Breen, John P
;      REGISTRATION NUMBER:  38,833
;      REFERENCE/DOCKET NUMBER:  835910-28685
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 231-7745
;      TELEFAX:  (317) 231-7433
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5433 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium yoelii
US-08-929-329-1

  Query Match             1.1%;  Score 44.8;  DB 3;  Length 5433;
  Best Local Similarity   46.2%;  Pred. No. 0.11;
  Matches  148;  Conservative    0;  Mismatches  172;  Indels    0;  Gaps    0;

Qy       3724 AAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGATGAA 3783
              ||| |||  |    |   ||| | ||| |||  |||   ||   ||   |||  ||||
Db       3305 AAAATTAACAGTTATACACCAAATAGGAGGGGAGAAAATTTTGCAAAGGAAAGTGATTCT 3364

Qy       3784 CAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAG 3843
                        | || ||| ||    |  | || |  |||    ||  |||||| |
Db       3365 ACAAGAAATACCGATGAATCGAAAATGGATGAGGTGATAAGGAAACGTGAAGAAGCTGCA 3424

Qy       3844 AATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACA 3903
              || | | |     | ||   | |||   |    | ||| |||  ||    | | |   | |
Db       3425 AAGAATGCTGAGATAATAAGAAAATTTGAAGAAGCACAAAAGGCTGCGTGGGCAAAAAAA 3484

Qy       3904 CCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCT 3963
               || |||| ||    | || |       ||  ||||  |    |  | ||      ||
Db       3485 GCAGAAGAGGAAAGGAAAAAGGCTGAAGCTGTAAAAAAAGCAGAGGAGGAAAGAAAACGA 3544

Qy       3964 CCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCC 4023
               | | |  ||  || | | ||| || |||||| || |      |  | ||| |||| |
Db       3545 ATTGAAGCTGAAAAGAAAGCAGAGGAGGAAAGAAAACGAATTGAAGCTGAAAAGAAAGCA 3604

Qy       4024 AAAAGCAAACAGCACCGAGT 4043
                ||    ||    | ||| |
Db       3605 GAAGAAGAAAGAAAACGAAT 3624

RESULT 7
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

  Query Match             1.1%;  Score 44;  DB 3;  Length 289;
  Best Local Similarity   10.4%;  Pred. No. 0.036;
  Matches   30;  Conservative  112;  Mismatches  146;  Indels    0;  Gaps    0;

Qy       3669 AGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAACAGAAACC 3728
              :|:   : : :|:|:  :: : ::: :::  :  :| |:  ::  |   : :|  : :
Db          1 RGRGRGRARCRARARURURARCRURARURURURARCRARARURURARCRARARURGRNRN 60

Qy       3729 TAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGATGAACAGCA 3788
               : :   :   ::: :  : : : : : : : : : : : :::   :::   :: : :::
Db         61 RSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRN 120

Qy       3789 GTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAG 3848
               : :    :: : : : : :: : :::: :      ::: :  :: : :  ::   : : :
Db        121 RSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRN 180

Qy       3849 TCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAA 3908
               :    : : : ::: : : : :  ::   ::  : ::     |: : :|: : :  |:|
Db        181 RSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRCRARGRCRURGRCRGRURARA 240

Qy       3909 AGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATC 3956
               : : : :    : :  |: :        |||||| ||  ||||||||| |
Db        241 RCRURCRURURGRCRGRCRUAAAAAAAAAAAAAAAAAAAAAAAAAAAC 288

RESULT 8
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
```

```
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

  Query Match             1.1%;  Score 44;  DB 3;  Length 289;
  Best Local Similarity   10.4%;  Pred. No. 0.036;
  Matches   30;  Conservative  112;  Mismatches  146;  Indels    0;  Gaps    0;

Qy       3669 AGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAACAGAAACC 3728
              :|:   : : :|:|:  :: : ::: :::  :  :| |:  ::  |   : :|   : :
Db          1 RGRGRGRARCRARARURURARCRURARURURURARCRARARURURARCRARARURGRNRN 60

Qy       3729 TAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGATGAACAGCA 3788
                : :   :    ::: :  :: : : : :  :: : ::     :::     :: : :::
Db         61 RSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRN 120

Qy       3789 GTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAG 3848
              : :    :: : : : : ::  : ::: :          ::: :  :: :  ::   : : :
Db        121 RSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRN 180

Qy       3849 TCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAA 3908
                :     : : : ::: : : : :  ::    ::  : ::      |: : :|: : :  |:|
Db        181 RSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRCRARGRCRURGRCRGRURARA 240

Qy       3909 AGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATC 3956
              : : : :    : :  |: :       ||||||  ||  |||||||||  |
Db        241 RCRURCRURURGRCRGRCRUAAAAAAAAAAAAAAAAAAAAAAAAAAAAAC 288

RESULT 9
US-09-134-001C-1924/c
; Sequence 1924, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1924
;   LENGTH: 435
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1924

  Query Match             1.0%;  Score 43.6;  DB 4;  Length 435;
  Best Local Similarity   49.1%;  Pred. No. 0.058;
  Matches  115;  Conservative    0;  Mismatches  119;  Indels    0;  Gaps    0;

Qy       3796 GAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCA 3855
              ||    |||  |   || ||  |||||| |    || || |||  | || |  |   |   | |||
Db        380 GACACAAAAACACTAAAGAAAGATAAAGATGAGAAAGAATCTAAAGACACAAAAACACCA 321

Qy       3856 AAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAG 3915
              || || |   |||  |       |       |||| |||          || || |   | |   || |
Db        320 AAGAAAGACAAAGAAAAGAAAGACATAAAAACTCCGAAGAAAGATAGAGAAGAGAAAAAA 261

Qy       3916 CCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAG 3975
              |||  |||    || |         ||||       ||| |  ||          |     || ||   | |
Db        260 CCAGTAATACCAAGTAACGGTAAAGTCGAGAAAGACGAAGACACAAAAACACCAAAAAAA 201

Qy       3976 GAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGC 4029
              ||   | ||| |   ||    | | ||| |    || |   ||          ||    ||||||
Db        200 GACAAAGAAAAGAAAATCACAAAAACTTTGAAGAAAGATAGAGAGGAGAAAAGC 147

RESULT 10
US-09-134-001C-1833
; Sequence 1833, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1833
;   LENGTH: 1089
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1833

  Query Match             1.0%;  Score 43.6;  DB 4;  Length 1089;
  Best Local Similarity   49.1%;  Pred. No. 0.095;
  Matches  115;  Conservative    0;  Mismatches  119;  Indels    0;  Gaps    0;

Qy       3796 GAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCA 3855
              ||    |||  |   || ||  |||||| |    || || |||  | || |  |   |   | |||
Db        634 GACACAAAAACACTAAAGAAAGATAAAGATGAGAAAGAATCTAAAGACACAAAAACACCA 693

Qy       3856 AAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAG 3915
              || || |   |||  |       |       |||| |||          || || |   | |   || |
Db        694 AAGAAAGACAAAGAAAAGAAAGACATAAAAACTCCGAAGAAAGATAGAGAAGAGAAAAAA 753

Qy       3916 CCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAG 3975
              |||  |||    || |         ||||       ||| |  ||          |     || ||   | |
Db        754 CCAGTAATACCAAGTAACGGTAAAGTCGAGAAAGACGAAGACACAAAAACACCAAAAAAA 813

Qy       3976 GAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGC 4029
              ||   | ||| |   ||    | | ||| |    || |   ||          ||    ||||||
Db        814 GACAAAGAAAAGAAAATCACAAAAACTTTGAAGAAAGATAGAGAGGAGAAAAGC 867

RESULT 11
US-08-155-888-1
; Sequence 1, Application US/08155888
; Patent No. 6066623
;  GENERAL INFORMATION:
;    APPLICANT:  Hoffman, Stephen L.
;    APPLICANT:  Hedstrom, Richard C.
;    APPLICANT:  Sedegah, Martha
;    TITLE OF INVENTION:  POLYNUCLEOTIDE VACCINE PROTECTIVE
```

```
;     TITLE OF INVENTION:  AGAINST MALARIA, METHODS OF PROTECTION AND VECTOR FOR
;     TITLE OF INVENTION:  DELIVERING POLYNUCLEOTIDE VACCINES
;     NUMBER OF SEQUENCES:  2
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Naval Medical Res. & Dev. Cmd.
;       STREET:  Bldg. 1, T-12 8901 Wisconsin Ave.
;       CITY:  Bethesda
;       STATE:  Maryland
;       COUNTRY:  USA
;       ZIP:  20889-5606
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/155,888
;       FILING DATE:
;       CLASSIFICATION:  514
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Spevack, A. David
;       REGISTRATION NUMBER:  24,743
;       REFERENCE/DOCKET NUMBER:  N.C. 75,851
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 295-6759
;       TELEFAX:  (202) 295-1022
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5552 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  circular
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FEATURE:
;       NAME/KEY:  promoter
;       LOCATION:  1..755
;       IDENTIFICATION METHOD:  experimental
;       OTHER INFORMATION:  /function= "promoter"
;       OTHER INFORMATION:  /evidence= EXPERIMENTAL
;       OTHER INFORMATION:  /label= CMV-IE
;       OTHER INFORMATION:  /note= "This feature acts as a promoter for any
;       OTHER INFORMATION:  downstream DNA sequence."
;       OTHER INFORMATION:  /citation= ([2])
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  933..2367
;       IDENTIFICATION METHOD:  experimental
;       OTHER INFORMATION:  /codon_start= 933
;       OTHER INFORMATION:  /function= "protein protective against malaria"
;       OTHER INFORMATION:  /product= "protein"
;       OTHER INFORMATION:  /evidence= EXPERIMENTAL
;       OTHER INFORMATION:  /number= 1
;       OTHER INFORMATION:  /label= IL2-CSP
;       OTHER INFORMATION:  /citation= ([1])
;     PUBLICATION INFORMATION:
;       AUTHORS:  Sedegah, Martha
;       AUTHORS:  Hedstrom, Richard C.
;       AUTHORS:  Hoffman, Stephen L.
;       TITLE:  Vaccination with Plasmodium yoelii CS protein
;       TITLE:  plasmid DNA protects against malaria
;       JOURNAL:  Science
;     PUBLICATION INFORMATION:
;       AUTHORS:  Cullen, Bryan R.
;       TITLE:  TRANS-ACTIVATION OF HUMAN IMMUNODEFICIENCY
;       TITLE:  VIRUS OCCURS VIA A BIMODAL MECHANISM
;       JOURNAL:  CELL
;       VOLUME:  46
;       PAGES:  973-982
;       DATE:  26 SEP-1986
;       RELEVANT RESIDUES IN SEQ ID NO:  1: FROM 1 TO 4732
US-08-155-888-1

  Query Match             1.0%;  Score 43.6;  DB 3;  Length 5552;
  Best Local Similarity   49.1%;  Pred. No. 0.23;
  Matches  115;  Conservative    0;  Mismatches  119;  Indels    0;  Gaps    0;

Qy       3771 AGAATCTGATGAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAA 3830
Db       1415 AAAACCAGAAGAAAAAAAAGATGATCCCCCAAAAGATGGCAACAAAGATGATCTTCCAAA 1474

Qy       3831 TGAAGATGAACAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCT 3890
Db       1475 AGAAGAAAAAAAAGATGATCTTCCAAAAGAAGAAAAAAAAGATGATCCCCCAAAGATCC 1534

Qy       3891 TGGTGGAGGTACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAA 3950
Db       1535 TAAAAAAGATGATCCACCAAAAGAGGCTCAAAATAAATTGAATCAACCAGTAGTGGCAGA 1594

Qy       3951 AAAATCTGGACCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGG 4004
Db       1595 TGAAAATGTAGATCAAGGGCCAGGAGCACCACAAGGGCCAGGAGCACCACAAGG 1648


RESULT 12
US-09-434-408-3
; Sequence 3, Application US/09434408
; Patent No. 6440697
; GENERAL INFORMATION:
;  APPLICANT: Venezia, Domenick
;  APPLICANT: Grossmann, Angelika
;  TITLE OF INVENTION: RING FINGER PROTEIN ZAPOP3
;  FILE REFERENCE: 98-41
;  CURRENT APPLICATION NUMBER: US/09/434,408
;  CURRENT FILING DATE: 1999-11-04
;  EARLIER APPLICATION NUMBER: US 60/108,258
;  EARLIER FILING DATE: 1998-11-12
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 2169
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Degenerate polynucleotide sequence of zapop3
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(2169)
;   OTHER INFORMATION: n = A,T,C or G
US-09-434-408-3

  Query Match             1.0%;  Score 42.4;  DB 4;  Length 2169;
  Best Local Similarity   23.6%;  Pred. No. 0.29;
  Matches   76;  Conservative   73;  Mismatches  173;  Indels    0;  Gaps    0;

Qy       3779 ATGAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATG 3838
Db        842 AYACNCARYTNYTNCARCARWSNWSNWSNCARAARGAYGARATHYTNCARACNGTNAARG 901

Qy       3839 AACAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAG 3898
Db        902 ARGARCARWSNMGNYTNGARCARGGNYTNWSNGARCAYCARMGNCAYYTNGAYGCNGARM 961

Qy       3899 GTACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTG 3958
Db        962 GNCARMGNYTNCARGARCARYTNAARCARACNGARCARAAYATHWSNWSNMGNATHCARA 1021

Qy       3959 GACCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGA 4018
Db       1022 ARYTNYTNCARGAYAAYCARMGNCARAARAARWSNWSNGARATHYTNAARWSNYTNGARA 1081

Qy       4019 AGTCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTG 4078
Db       1082 AYGARMGNATHMGNATGGARCARYTNATGWSNATHACNCARGARGARACNGARWSNYTNM 1141

Qy       4079 AATCTAGTGCAATTGAATCCAC 4100
```

```
                 :| |   | |  ::  |
Db       1142 GNMGNMGNGAYGTNGCNWSNGC 1163

RESULT 13
PCT-US93-07261-10
; Sequence 10, Application PC/TUS9307261
;  GENERAL INFORMATION:
;    TITLE OF INVENTION:  PfEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John H. C. Blasdale
;      STREET:  One Giralda Farms
;      CITY:  Madison
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07940-1000
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 6.0.5
;      SOFTWARE:  Microsoft Word 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/07261
;      FILING DATE:  19930805
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/927,531
;      FILING DATE:  07-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Blasdale, John H. C.
;      REGISTRATION NUMBER:  31,895
;      REFERENCE/DOCKET NUMBER:  DX0288K
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-822-7398
;      TELEFAX:  201-822-7039
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4766 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium falciparum
;      STRAIN:  Malayan Camp
;    IMMEDIATE SOURCE:
;      CLONE:  p2b1;p12-1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..4766
PCT-US93-07261-10

  Query Match             1.0%;  Score 42.4;  DB 5;  Length 4766;
  Best Local Similarity   48.7%;  Pred. No. 0.44;
  Matches  115;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy       3797 AGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCAA 3856
              |||||||| | | |||||| | | |    | | |   | ||   ||   |     |||
Db         67 AGGAAAAGAAGTTAAAAGAACAAAAAAAGAATGACGCGCAGAAAGCAAAAGATTTAACAA 126

Qy       3857 AAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAGC 3916
              ||||||  |     || |    ||    ||          |  |  | ||| | |||
Db        127 AAAAGGAAAGCCAAGACTCTTCAAGTGAAAAATCACTAAAGGAAAAAGTAAATGGAGAAG 186

Qy       3917 CAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAAATCTGGACCTCCAGCACCAGAGG 3976
              || || | |||     || |||| ||    |||||||  |          |   |   |||
Db        187 CATTAAAGGAAAAAGAAAATAAAGAAACATTAAAAAAAAAAGAGTTAGAAAATCAGAAGG 246

Qy       3977 AGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAA 4032
              |  | |||||| || | | |   ||  |  ||||  ||     |   || |  |||
Db        247 AAAAAGAAGAAAAAAAATAAAATAAAAGATAATAATGATGAAGCATTAAAGAATAAA 302
```

```
RESULT 14
US-08-948-705-3
; Sequence 3, Application US/08948705A
; Patent No. 6043084
; GENERAL INFORMATION:
;  APPLICANT: Scanlan, Matthew J.
;  APPLICANT: Chen, Yao-Tseng
;  APPLICANT: Stockert, Elisabeth
;  APPLICANT: Old, Lloyd J.
;  TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES
;  TITLE OF INVENTION: ASSOCIATED WITH COLON CANCER AND METHODS FOR DIAGNOSING AND
;  TITLE OF INVENTION: TREATING COLON CANCER
;  FILE REFERENCE: LUD-5506-JEL/NDH
;  CURRENT APPLICATION NUMBER: US/08/948,705A
;  CURRENT FILING DATE: 1997-10-10
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 1298
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 55..55
;   OTHER INFORMATION:
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 141..141
;   OTHER INFORMATION:
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 199..99
;   OTHER INFORMATION:
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 342..342
;   OTHER INFORMATION:
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 352..352
;   OTHER INFORMATION:
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 722..722
;   OTHER INFORMATION:
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 750..750
;   OTHER INFORMATION:
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 1058..1058
;   OTHER INFORMATION:
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 1101..1101
;   OTHER INFORMATION:
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 1144..1144
;   OTHER INFORMATION:
US-08-948-705-3

  Query Match             1.0%;  Score 42;  DB 3;  Length 1298;
  Best Local Similarity   49.4%;  Pred. No. 0.28;
  Matches  131;  Conservative    0;  Mismatches  133;  Indels    1;  Gaps    1;

Qy       3801 AAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCAAAAAA 3860
              ||||| |  | |||||||    |    ||| |  |||| | | |||| |        ||   |
Db       1000 AAAGAAGAACTAAGAAGAAGGAGAGGAAGAATAAGAAGGAAAGAAGAAAGAAAAAAGTNA 1059

Qy       3861 GGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAGCCAAC 3920
                | | ||||        |  |  |        |          |  | |||||     |
```

```
Db       1060 AAGAAGAAGAAAGAAGGAAGAAGGAAAGAAGAGGAAGAACTNAGAAGAAGAAAGAGGAGG 1119

Qy       3921 AATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGGA 3980

Db       1120 AAAGAAGAAAGAAGAATAAGGAACNAGAAAGAAGGAGAAGAAAGAATAAGAAGAGGAAGA 1179

Qy       3981 GGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAACAGCACCG 4040

Db       1180 AGAAAAAGAAGAAA-AGAAGAAGGAAAGAAGGAGAAAAAGGAAGAAAAAAGGAAGAAGAA 1238

Qy       4041 AGTGTCAAGGAGAGCACAGCAGAGA 4065

Db       1239 AGTAGAAAGCGGAAGAAAGAAAAGA 1263


RESULT 15
US-09-510-543-3
; Sequence 3, Application US/09510543
; Patent No. 6517837
; GENERAL INFORMATION:
;  APPLICANT:  Scanlan, Matthew J.
;  APPLICANT:  Chen, Yao-Tseng
;  APPLICANT:  Stockert, Elisabeth
;  APPLICANT:  Old, Lloyd J.
;  TITLE OF INVENTION:  ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED WITH COLON CANCER AND
;  TITLE OF INVENTION:  METHODS FOR DIAGNOSING AND TREATING COLON CANCER
;  FILE REFERENCE:  LUD-5506.1-JEL/NDH
;  CURRENT APPLICATION NUMBER: US/09/510,543
;  CURRENT FILING DATE:  2000-02-22
;  PRIOR APPLICATION NUMBER:  US 09/102,322
;  PRIOR FILING DATE:  1998-06-22
;  NUMBER OF SEQ ID NOS:  22
; SEQ ID NO 3
;   LENGTH: 1298
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 55..55
;   OTHER INFORMATION:
;   NAME/KEY: unsure
;   LOCATION: 141..141
;   OTHER INFORMATION:
;   NAME/KEY: unsure
;   LOCATION: 199..199
;   OTHER INFORMATION:
;   NAME/KEY: unsure
;   LOCATION: 342..342
;   OTHER INFORMATION:
;   NAME/KEY: unsure
;   LOCATION: 352..352
;   OTHER INFORMATION:
;   NAME/KEY: unsure
;   LOCATION: 722..722
;   OTHER INFORMATION:
;   NAME/KEY: unsure
;   LOCATION: 750..750
;   OTHER INFORMATION:
;   NAME/KEY: unsure
;   LOCATION: 1058..1058
;   OTHER INFORMATION:
;   NAME/KEY: unsure
;   LOCATION: 1101..1101
;   OTHER INFORMATION:
;   NAME/KEY: unsure
;   LOCATION: 1144..1144
;   OTHER INFORMATION:
US-09-510-543-3

  Query Match             1.0%;  Score 42;  DB 4;  Length 1298;
  Best Local Similarity   49.4%;  Pred. No. 0.28;
  Matches  131;  Conservative    0;  Mismatches  133;  Indels    1;  Gaps    1;

Qy       3801 AAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCAAAAAA 3860

Db       1000 AAAGAAGAACTAAGAAGAAGGAGAGGAAGAATAAGAAGGAAAGAAGAAAGAAAAAAGTNA 1059

Qy       3861 GGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAGCCAAC 3920

Db       1060 AAGAAGAAGAAAGAAGGAAGAAGGAAAGAAGAGGAAGAACTNAGAAGAAGAAAGAGGAGG 1119

Qy       3921 AATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGGA 3980

Db       1120 AAAGAAGAAAGAAGAATAAGGAACNAGAAAGAAGGAGAAGAAAGAATAAGAAGAGGAAGA 1179

Qy       3981 GGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAACAGCACCG 4040

Db       1180 AGAAAAAGAAGAAA-AGAAGAAGGAAAGAAGGAGAAAAAGGAAGAAAAAAGGAAGAAGAA 1238

Qy       4041 AGTGTCAAGGAGAGCACAGCAGAGA 4065

Db       1239 AGTAGAAAGCGGAAGAAAGAAAAGA 1263

Search completed: September 24, 2003, 16:31:55
Job time : 227 secs
```

```
                 GenCore version 5.1.6
        Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        September 24, 2003, 07:23:42 ; Search time 1011 Seconds
                                        (without alignments)
                                        11142.191 Million cell updates/sec

Title:         US-09-512-581B-3
Perfect score: 4173
Sequence:      1 atggctcattcaaagactag..........cacaaccaaaaaaaaatgtg 4173

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters:      5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       N_Geneseq_19Jun03:*
               1:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
               2:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
               3:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
               4:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
               5:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
               6:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
               7:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
               8:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
               9:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
               10:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
               11:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
               12:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
               13:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
               14:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
               15:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
               16:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
               17:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
               18:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
               19:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
               20:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
               21:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
               22:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
               23:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
               24:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
               25:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4173	100.0	5271	21	AAA28051	Human androgen shu
	2	4173	100.0	5355	21	AAA28052	Human androgen shu
	3	4165.6	99.8	7473	23	ABV22430	Human prostate exp
	4	4164	99.8	7473	23	ABV25469	Human prostate exp
	5	4164	99.8	7473	23	ABV28244	Human prostate exp
	6	1815.6	43.5	6744	25	ABX14056	cDNA encoding huma
	7	1157.2	27.7	5469	24	ABQ99271	Human coding seque
	8	1112.8	26.7	3957	21	AAA47423	Sequence encoding
	9	499	12.0	714	22	AAS44727	Human full-length
	10	493	11.8	2496	24	ABK35358	Human cDNA encodin
	11	438	10.5	7943	22	AAK85948	Human immune/haema
	12	411.4	9.9	439	20	AAV87629	EST clone DY17. H
	13	353.4	8.5	461	24	ABL89669	Human polynucleoti
	14	350.4	8.4	1104	24	ABS56684	Euchromosome fragi
	15	348	8.3	738	20	AAZ15259	Human gene express
	16	308	7.4	3920	23	ABL22105	Drosophila melanog
	17	298.8	7.2	351	21	AAZ42861	Human 5' EST isola
	18	292	7.0	295	21	AAC02817	Human secreted pro
	19	253.8	6.1	1217	21	AAF18338	Lung cancer associ
	20	244.2	5.9	2295	22	AAS02396	Human secreted pro
	21	232.2	5.6	772	22	AAH03894	Human cDNA clone (
	22	232.2	5.6	2212	22	AAH17132	Human cDNA sequenc
	23	208.8	5.0	227	24	ABQ94421	Tumour suppression
c	24	203.6	4.9	530	21	AAZ80598	Human colon cancer
	25	164	3.9	433	22	AAS44899	Human contig polyn
c	26	119.2	2.9	6990	23	ABL22104	Drosophila melanog
	27	108	2.6	785	22	AAH07766	Human cDNA clone (
	28	108	2.6	1489	22	AAH18066	Human cDNA sequenc
	29	99	2.4	161	21	AAC06611	Human secreted pro
c	30	94	2.3	557	24	ABK53631	Human eosinophil-m
	31	87.4	2.1	331	24	ABS69109	Novel murine polyn
c	32	64.4	1.5	14006	24	ABL33958	Human immune syste
	33	62.8	1.5	1406	23	ABV22097	Human prostate exp
	34	62.8	1.5	1406	23	ABV23050	Human prostate exp
	35	62.8	1.5	1406	23	ABV27937	Human prostate exp
	36	62.8	1.5	1406	23	ABV28886	Human prostate exp
c	37	62.8	1.5	6668	24	ABL33697	Human immune syste
c	38	62.2	1.5	464	23	ABV55326	Human prostate exp
c	39	62.2	1.5	469	23	ABV55197	Human prostate exp
c	40	62.2	1.5	556	24	ABQ36996	Oligonucleotide fo
	41	62.2	1.5	556	24	ABQ36997	Oligonucleotide fo
c	42	62.2	1.5	34769	22	AAS46774	Tumour suppressor
c	43	61.2	1.5	887	22	AAI94064	Human neuroblastom
	44	61.2	1.5	1267	24	ABZ78031	Human breast speci
c	45	60.6	1.5	9539	22	AAS45347	Chemically pretrea

ALIGNMENTS

```
RESULT 1
AAA28051
ID   AAA28051 standard; cDNA; 5271 BP.
XX
AC   AAA28051;
XX
DT   01-DEC-2000  (first entry)
XX
DE   Human androgen shutoff gene 3 (AS3) cDNA sequence.
XX
KW   Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;
KW   chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;
KW   diagnosis; treatment; cytostatic; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   exon            1..46
FT                   /*tag= a
FT                   /number= 1
FT   exon            47..173
FT                   /*tag= b
FT                   /number= 2
FT   CDS             66..4241
FT                   /*tag= c
FT                   /product= "AS3"
FT                   /note= "Androgen shutoff gene 3 protein, the CDS is
FT                          specifically claimed as SEQ ID #3"
FT   exon            174..377
FT                   /*tag= d
FT                   /number= 3
```

```
FT   exon            378..464
FT                   /*tag= e
FT                   /number= 4
FT   exon            465..562
FT                   /*tag= f
FT                   /number= 5
FT   exon            563..689
FT                   /*tag= e
FT                   /number= 6
FT   exon            690..854
FT                   /*tag= f
FT                   /number= 7
FT   exon            855..911
FT                   /*tag= g
FT                   /number= 8
FT   exon            912..1027
FT                   /*tag= h
FT                   /number= 9
FT   exon            1028..1122
FT                   /*tag= i
FT                   /number= 10
FT   exon            1123..1268
FT                   /*tag= j
FT                   /number= 11
FT   exon            1269..1420
FT                   /*tag= k
FT                   /number= 12
FT   exon            1421..1534
FT                   /*tag= l
FT                   /number= 13
FT   exon            1535..1616
FT                   /*tag= m
FT                   /number= 14
FT   exon            1617..1665
FT                   /*tag= n
FT                   /number= 15
FT   exon            1666..1805
FT                   /*tag= o
FT                   /number= 16
FT   exon            1806..1921
FT                   /*tag= p
FT                   /number= 17
FT   exon            1922..2027
FT                   /*tag= q
FT                   /number= 18
FT   exon            2028..2188
FT                   /*tag= r
FT                   /number= 19
FT   exon            2189..2312
FT                   /*tag= s
FT                   /number= 20
FT   exon            2313..2471
FT                   /*tag= t
FT                   /number= 21
FT   exon            2472..2540
FT                   /*tag= u
FT                   /number= 22
FT   exon            2541..2677
FT                   /*tag= v
FT                   /number= 23
FT   exon            2678..2801
FT                   /*tag= w
FT                   /number= 24
FT   exon            2802..3006
FT                   /*tag= x
FT                   /number= 25
FT   exon            3007..3121
FT                   /*tag= y
FT                   /number= 26
FT   exon            3122..3254
FT                   /*tag= z
FT                   /number= 27
FT   exon            3255..3374
FT                   /*tag= aa
FT                   /number= 28
FT   exon            3375..3437
FT                   /*tag= ab
FT                   /number= 29
FT   exon            3438..3583
FT                   /*tag= ac
FT                   /number= 30
FT   exon            3584..3689
FT                   /*tag= ad
FT                   /number= 31
FT   exon            3690..4129
FT                   /*tag= ae
FT                   /number= 32
FT   exon            4130..4354
FT                   /*tag= af
FT                   /number= 33
FT   misc_signal     4320..4324
FT                   /*tag= ag
FT                   /note= "Destabilizing signal"
FT   exon            4355..5253
FT                   /*tag= ah
FT                   /number= 34
FT   polyA_signal    5228..5233
FT                   /*tag= ai
FT   misc_feature    5249..5253
FT                   /*tag= aj
FT                   /note= "Poly adenylation cleavage site"
XX
PN   WO200050454-A1.
XX
PD   31-AUG-2000.
XX
PF   24-FEB-2000; 2000WO-US04732.
XX
PR   24-FEB-1999;   99US-0121461.
XX
PA   (TUFT ) TUFTS COLLEGE.
XX
PI   Soto AM,  Sonnenschein C,  Geck P,  Szelei J;
XX
DR   WPI; 2000-565451/52.
DR   P-PSDB; AAY94702.
XX
PT   New human androgen-induced tumor suppressor cDNA sequence termed
PT   'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
PT   diagnosis and treatment of prostate cancer -
XX
PS   Claim 1; Fig 1; 152pp; English.
XX
CC   This invention relates to a human androgen-induced tumour suppressor cDNA
CC   sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is located
CC   on chromosome 13 at position 13q12-13q. AS3 has a role in inhibiting cell
CC   proliferation and use as a marker for the efficient diagnosis and
CC   treatment of prostate cancer. The invention includes AS3 cDNA and protein
CC   sequences, a vector comprising the cDNA sequence, a host cell transfected
CC   with the expression vector, and a method for producing an AS3 polypeptide
CC   comprising culturing the transfected cells. AS3 has cytostatic activity,
CC   and acts to suppress cell proliferation. The AS3 gene is useful as a
CC   marker for the efficient diagnosis and treatment of prostate cancer. The
CC   AS3 nucleic acid molecule can be used as a source of antisense agents for
CC   sequence specific modulation of gene expression. The AS3 protein may be
CC   used in the treatment of disorders caused by aberrant modification or
CC   mutation of a gene encoding an AS3 protein, misregulation of the AS3 gene
CC   or aberrant post-translational modification of the AS3 protein. This
CC   sequence represents the human AS3 cDNA sequence.
XX
SQ   Sequence 5271 BP; 1782 A; 944 C; 1066 G; 1479 T; 0 other;

  Query Match             100.0%;  Score 4173;  DB 21;  Length 5271;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4173;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 125

Qy         61 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 185

Qy        121 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 245

Qy        181 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGGTAAAGATGTTCGCTTACTGGTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGGTAAAGATGTTCGCTTACTGGTA 305

Qy        241 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 365

Qy        301 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 425

Qy        361 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 485

Qy        421 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 545

Qy        481 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 605

Qy        541 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 665

Qy        601 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 725

Qy        661 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCACTTTTTTTAATCAGGTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCACTTTTTTTAATCAGGTT 785

Qy        721 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        786 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 845

Qy        781 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 905

Qy        841 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 965

Qy        901 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 1025

Qy        961 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1085

Qy       1021 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1145

Qy       1081 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1205

Qy       1141 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1206 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1265

Qy       1201 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1266 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1325

Qy       1261 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1326 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1385

Qy       1321 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1386 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1445

Qy       1381 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1446 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1505

Qy       1441 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1506 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1565

Qy       1501 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1566 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1625

Qy       1561 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1626 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1685

Qy       1621 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1686 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1745

Qy       1681 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1746 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1805

Qy       1741 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1806 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1865

Qy       1801 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1866 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 1925

Qy       1861 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1926 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 1985

Qy       1921 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1986 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 2045

Qy       1981 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2046 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2105

Qy       2041 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2106 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2165

Qy       2101 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2166 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2225

Qy       2161 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTCTAGTAAA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2226 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTCTAGTAAA 2285

Qy       2221 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2286 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2345

Qy       2281 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2346 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2405

Qy       2341 TTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2406 TTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2465

Qy       2401 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2466 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2525

Qy       2461 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2526 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2585

Qy       2521 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2586 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2645

Qy       2581 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2646 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2705

Qy       2641 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2706 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2765

Qy       2701 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2766 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2825

Qy       2761 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2826 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2885

Qy       2821 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2886 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 2945

Qy       2881 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2946 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 3005

Qy       2941 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3006 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3065

Qy       3001 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3066 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3125

Qy       3061 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3126 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3185

Qy       3121 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3186 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3245

Qy       3181 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3246 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3305

Qy       3241 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3306 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3365

Qy       3301 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3366 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3425

Qy       3361 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3426 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3485

Qy       3421 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3486 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3545

Qy       3481 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3546 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3605

Qy       3541 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3606 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3665

Qy       3601 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3666 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3725

Qy       3661 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3726 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3785

Qy       3721 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3786 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3845

Qy       3781 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3846 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3905

Qy       3841 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3906 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 3965

Qy       3901 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 4025

Qy       3961 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4026 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4085

Qy       4021 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4086 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4145

Qy       4081 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4146 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4205

Qy       4141 ACGCCATCACCATCACAACCAAAAAAAAATGTG 4173
              |||||||||||||||||||||||||||||||||
Db       4206 ACGCCATCACCATCACAACCAAAAAAAAATGTG 4238
```

```
RESULT 2
AAA28052
ID   AAA28052 standard; cDNA; 5355 BP.
XX
AC   AAA28052;
XX
DT   01-DEC-2000  (first entry)
XX
DE   Human androgen shutoff gene 3 (AS3) cDNA sequence SEQ ID #4.
XX
KW   Androgen-induced tumour suppressor; androgen shutoff gene 3; AS3;
```

```
KW   chromosome 13q12-13q; cell proliferation inhibitor; prostate cancer;
KW   diagnosis; treatment; cytostatic; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   exon            85..130
FT                   /*tag=  a
FT                   /number= 1
FT   exon            131..257
FT                   /*tag=  b
FT                   /number= 2
FT   CDS             150..4325
FT                   /*tag=  c
FT                   /product= "AS3"
FT                   /note= "Androgen shutoff gene 3 protein"
FT   exon            258..461
FT                   /*tag=  d
FT                   /number= 3
FT   exon            462..548
FT                   /*tag=  e
FT                   /number= 4
FT   exon            549..646
FT                   /*tag=  f
FT                   /number= 5
FT   exon            647..773
FT                   /*tag=  e
FT                   /number= 6
FT   exon            774..938
FT                   /*tag=  f
FT                   /number= 7
FT   exon            939..995
FT                   /*tag=  g
FT                   /number= 8
FT   exon            996..1111
FT                   /*tag=  h
FT                   /number= 9
FT   exon            1112..1206
FT                   /*tag=  i
FT                   /number= 10
FT   exon            1207..1352
FT                   /*tag=  j
FT                   /number= 11
FT   exon            1353..1504
FT                   /*tag=  k
FT                   /number= 12
FT   exon            1505..1618
FT                   /*tag=  l
FT                   /number= 13
FT   exon            1619..1700
FT                   /*tag=  m
FT                   /number= 14
FT   exon            1701..1749
FT                   /*tag=  n
FT                   /number= 15
FT   exon            1750..1889
FT                   /*tag=  o
FT                   /number= 16
FT   exon            1890..2005
FT                   /*tag=  p
FT                   /number= 17
FT   exon            2006..2111
FT                   /*tag=  q
FT                   /number= 18
FT   exon            2112..2272
FT                   /*tag=  r
FT                   /number= 19
FT   exon            2273..2396
FT                   /*tag=  s
FT                   /number= 20
FT   exon            2397..2555
FT                   /*tag=  t
FT                   /number= 21
FT   exon            2556..2624
FT                   /*tag=  u
FT                   /number= 22
FT   exon            2625..2761
FT                   /*tag=  v
FT                   /number= 23
FT   exon            2672..2885
FT                   /*tag=  w
FT                   /number= 24
FT   exon            2886..3090
FT                   /*tag=  x
FT                   /number= 25
FT   exon            3091..3205
FT                   /*tag=  y
FT                   /number= 26
FT   exon            3206..3338
FT                   /*tag=  z
FT                   /number= 27
FT   exon            3339..3458
FT                   /*tag=  aa
FT                   /number= 28
FT   exon            3459..3521
FT                   /*tag=  ab
FT                   /number= 29
FT   exon            3522..3667
FT                   /*tag=  ac
FT                   /number= 30
FT   exon            3668..3773
FT                   /*tag=  ad
FT                   /number= 31
FT   exon            3774..4213
FT                   /*tag=  ae
FT                   /number= 32
FT   exon            4214..4438
FT                   /*tag=  af
FT                   /number= 33
FT   misc_signal     4314..4408
FT                   /*tag=  ag
FT                   /note= "Destabilizing signal"
FT   exon            4439..5337
FT                   /*tag=  ah
FT                   /number= 34
FT   polyA_signal    5312..5317
FT                   /*tag=  ai
FT   misc_feature    5333..5337
FT                   /*tag=  aj
FT                   /note= "Poly adenylation cleavage site"
XX
PN   WO200050454-A1.
XX
PD   31-AUG-2000.
XX
PF   24-FEB-2000; 2000WO-US04732.
XX
PR   24-FEB-1999;   99US-0121461.
XX
PA   (TUFT ) TUFTS COLLEGE.
XX
PI   Soto AM,  Sonnenschein C,  Geck P,  Szelei J;
XX
DR   WPI; 2000-565451/52.
DR   P-PSDB; AAY94702.
XX
PT   New human androgen-induced tumor suppressor cDNA sequence termed
PT   'Androgen Shutoff Gene 3' (AS3), useful as a marker for the efficient
PT   diagnosis and treatment of prostate cancer -
XX
PS   Example 4; Fig 6; 152pp; English.
XX
CC   This invention relates to a human androgen-induced tumour suppressor
CC   cDNA sequence termed "Androgen Shutoff Gene 3" (AS3). The AS3 gene is
CC   located on chromosome 13 at position 13q12-13q. AS3 has a role in
CC   inhibiting cell proliferation and use as a marker for the efficient
```

```
CC   diagnosis and treatment of prostate cancer. The invention includes AS3
CC   cDNA and protein sequences, a vector comprising the cDNA sequence, a host
CC   cell transfected with the expression vector, and a method for producing
CC   an AS3 polypeptide comprising culturing the transfected cells. AS3 has
CC   cytostatic activity, and acts to suppress cell proliferation. The AS3
CC   gene is useful as a marker for the efficient diagnosis and treatment of
CC   prostate cancer. The AS3 nucleic acid molecule can be used as a source of
CC   antisense agents for sequence specific modulation of gene expression. The
CC   AS3 protein may be used in the treatment of disorders caused by aberrant
CC   modification or mutation of a gene encoding an AS3 protein, misregulation
CC   of the AS3 gene or aberrant post-translational modification of the AS3
CC   protein. This sequence represents the human AS3 cDNA sequence with an
CC   additional 84 nucleotides in the 5' untranslated region (5' UTR) when
CC   compared with the claimed AS3 cDNA sequence AAA28051.
XX
SQ   Sequence 5355 BP; 1798 A; 957 C; 1115 G; 1485 T; 0 other;

  Query Match             100.0%;  Score 4173;  DB 21;  Length 5355;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 4173;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        150 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 209

Qy         61 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        210 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 269

Qy        121 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        270 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 329

Qy        181 TTACATCTTGCTTCAGATTTTTTCTCAAGCATCCTGGTAAAGATGTTCGCTTACTGGTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        330 TTACATCTTGCTTCAGATTTTTTCTCAAGCATCCTGGTAAAGATGTTCGCTTACTGGTA 389

Qy        241 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        390 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 449

Qy        301 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        450 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 509

Qy        361 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        510 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 569

Qy        421 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        570 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 629

Qy        481 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        630 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 689

Qy        541 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        690 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 749

Qy        601 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        750 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 809

Qy        661 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCACTTTTTTAATCAGGTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        810 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCACTTTTTTAATCAGGTT 869

Qy        721 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        870 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 929

Qy        781 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        930 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 989

Qy        841 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        990 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 1049

Qy        901 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1050 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 1109

Qy        961 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1110 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1169

Qy       1021 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1170 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1229

Qy       1081 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1230 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1289

Qy       1141 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1290 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1349

Qy       1201 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1350 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1409

Qy       1261 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1410 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1469

Qy       1321 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1470 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1529

Qy       1381 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1530 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1589

Qy       1441 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1590 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1649

Qy       1501 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1650 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1709

Qy       1561 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1710 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1769

Qy       1621 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1770 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1829

Qy       1681 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1830 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1889

Qy       1741 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1890 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1949

Qy       1801 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1950 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 2009

Qy       1861 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 1920
```

```
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2010 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 2069

Qy        1921 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 1980
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2070 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 2129

Qy        1981 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2040
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2130 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2189

Qy        2041 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2190 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2249

Qy        2101 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2250 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2309

Qy        2161 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2310 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2369

Qy        2221 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2280
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2370 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2429

Qy        2281 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2340
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2430 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2489

Qy        2341 TTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2490 TTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2549

Qy        2401 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2460
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2550 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2609

Qy        2461 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2520
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2610 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2669

Qy        2521 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2580
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2670 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2729

Qy        2581 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2640
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2730 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2789

Qy        2641 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2700
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2790 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2849

Qy        2701 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2760
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2850 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2909

Qy        2761 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2820
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2910 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2969

Qy        2821 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 2880
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2970 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 3029

Qy        2881 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 2940
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3030 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 3089

Qy        2941 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3000
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3090 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3149

Qy        3001 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3060
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3150 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3209

Qy        3061 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3210 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3269

Qy        3121 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3270 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3329

Qy        3181 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3330 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3389

Qy        3241 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3390 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3449

Qy        3301 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3450 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3509

Qy        3361 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3510 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3569

Qy        3421 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3570 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3629

Qy        3481 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3630 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3689

Qy        3541 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3690 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3749

Qy        3601 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3750 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3809

Qy        3661 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3810 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3869

Qy        3721 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3870 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3929

Qy        3781 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3930 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3989

Qy        3841 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 3900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3990 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 4049

Qy        3901 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 3960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4050 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 4109

Qy        3961 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4110 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4169

Qy        4021 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4170 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4229
```

```
Qy       4081 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4230 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4289

Qy       4141 ACGCCATCACCATCACAACCAAAAAAAAATGTG 4173
              |||||||||||||||||||||||||||||||||
Db       4290 ACGCCATCACCATCACAACCAAAAAAAAATGTG 4322
```

RESULT 3
ABV22430

```
ID   ABV22430 standard; cDNA; 7473 BP.
XX
AC   ABV22430;
XX
DT   13-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 22421.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer  -
XX
PS   Claim 1; Page 3901-3902; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ   Sequence 7473 BP; 2458 A; 1243 C; 1480 G; 2285 T; 7 other;

  Query Match             99.8%;  Score 4165.6;  DB 23;  Length 7473;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 4168;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        158 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 217

Qy         61 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        218 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 277

Qy        121 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        278 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 337

Qy        181 TTACATCTTGCTTCAGATTTTTTCTCAAGCATCCTGGTAAAGATGTTCGCTTACTGGTA 240
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        338 TTACATCTTGCTTCAGATTTTTTCTCAAGCATCCTGATAAAGATGTTCGCTTACTGGTA 397

Qy        241 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        398 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 457

Qy        301 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        458 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 517

Qy        361 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        518 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 577

Qy        421 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        578 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 637

Qy        481 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        638 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 697

Qy        541 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        698 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 757

Qy        601 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        758 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 817

Qy        661 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCACTTTTTTAATCAGGTT 720
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        818 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTTTAATCAGGTT 877

Qy        721 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        878 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 937

Qy        781 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        938 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 997

Qy        841 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        998 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 1057

Qy        901 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1058 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 1117

Qy        961 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1118 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1177

Qy       1021 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1178 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1237

Qy       1081 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1238 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAG 1297

Qy       1141 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1298 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1357

Qy       1201 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1358 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1417

Qy       1261 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1418 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1477

Qy       1321 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1478 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1537

Qy       1381 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1538 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1597

Qy       1441 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1598 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1657

Qy       1501 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1658 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1717

Qy       1561 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1718 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1777

Qy       1621 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1778 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1837

Qy       1681 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1838 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1897

Qy       1741 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1898 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1957

Qy       1801 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1958 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 2017

Qy       1861 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2018 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 2077

Qy       1921 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2078 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 2137

Qy       1981 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2138 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2197

Qy       2041 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2198 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2257

Qy       2101 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2258 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2317

Qy       2161 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2318 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2377

Qy       2221 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2378 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2437

Qy       2281 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2438 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2497

Qy       2341 TTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2400
              ||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||
Db       2498 TTTGCTGCTCCTTTGAAATCTTTGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2557

Qy       2401 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2558 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2617

Qy       2461 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2618 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2677

Qy       2521 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2678 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2737

Qy       2581 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2738 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2797

Qy       2641 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2798 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2857

Qy       2701 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2858 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2917

Qy       2761 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2918 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2977

Qy       2821 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2978 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 3037

Qy       2881 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3038 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 3097

Qy       2941 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3098 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3157

Qy       3001 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3158 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3217

Qy       3061 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3218 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3277

Qy       3121 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3278 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3337

Qy       3181 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3338 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3397

Qy       3241 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3398 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3457
```

```
Qy       3301 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3458 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3517

Qy       3361 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3518 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3577

Qy       3421 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3578 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3637

Qy       3481 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3638 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3697

Qy       3541 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3698 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3757

Qy       3601 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3758 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3817

Qy       3661 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3818 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3877

Qy       3721 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3878 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3937

Qy       3781 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3938 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3997

Qy       3841 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3998 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 4057

Qy       3901 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4058 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 4117

Qy       3961 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4118 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4177

Qy       4021 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4178 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4237

Qy       4081 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4238 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4297

Qy       4141 ACGCCATCACCATCACAACCAAAAAAAAATGT 4172
              ||||||||||||||||||||||||||||||||
Db       4298 ACGCCATCACCATCACAACCAAAAAAAAATGT 4329
```

```
RESULT 4
ABV25469
ID   ABV25469 standard; cDNA; 7473 BP.
XX
AC   ABV25469;
XX
DT   16-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 25460.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer  -
XX
PS   Claim 1; Page 5055-5057; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ   Sequence 7473 BP; 2457 A; 1243 C; 1481 G; 2285 T; 7 other;

  Query Match             99.8%;  Score 4164;  DB 23;  Length 7473;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 4167;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        158 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 217

Qy         61 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        218 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 277

Qy        121 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        278 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 337

Qy        181 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGGTAAAGATGTTCGCTTACTGGTA 240
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        338 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGATAAAGATGTTCGCTTACTGGTA 397

Qy        241 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        398 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 457

Qy        301 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        458 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 517
```

```
Qy        361 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        518 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 577

Qy        421 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        578 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 637

Qy        481 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        638 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 697

Qy        541 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        698 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 757

Qy        601 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        758 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 817

Qy        661 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCACTTTTTTTAATCAGGTT 720
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        818 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTTTTAATCAGGTT 877

Qy        721 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        878 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 937

Qy        781 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        938 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 997

Qy        841 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        998 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 1057

Qy        901 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1058 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 1117

Qy        961 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1118 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1177

Qy       1021 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1178 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1237

Qy       1081 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1238 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1297

Qy       1141 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1298 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1357

Qy       1201 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1358 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1417

Qy       1261 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1418 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1477

Qy       1321 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1478 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1537

Qy       1381 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1538 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1597

Qy       1441 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1598 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1657

Qy       1501 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1658 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1717

Qy       1561 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1718 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1777

Qy       1621 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1680
              |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db       1778 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGGAGATGATGAGAAAATAAGA 1837

Qy       1681 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1838 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1897

Qy       1741 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1898 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1957

Qy       1801 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1958 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 2017

Qy       1861 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2018 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 2077

Qy       1921 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2078 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 2137

Qy       1981 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2138 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2197

Qy       2041 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2198 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2257

Qy       2101 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2258 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2317

Qy       2161 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2318 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2377

Qy       2221 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2378 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2437

Qy       2281 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2438 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2497

Qy       2341 TTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2400
              ||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||
Db       2498 TTTGCTGCTCCTTTGAAATCTTTGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2557

Qy       2401 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2558 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2617

Qy       2461 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2618 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2677

Qy       2521 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2580
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2678 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2737

Qy        2581 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2640
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2738 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2797

Qy        2641 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2700
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2798 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2857

Qy        2701 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2760
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2858 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2917

Qy        2761 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2820
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2918 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2977

Qy        2821 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 2880
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2978 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 3037

Qy        2881 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 2940
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3038 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 3097

Qy        2941 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3000
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3098 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3157

Qy        3001 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3060
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3158 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3217

Qy        3061 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3218 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3277

Qy        3121 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3278 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3337

Qy        3181 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3338 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3397

Qy        3241 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3398 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3457

Qy        3301 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3458 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3517

Qy        3361 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3518 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3577

Qy        3421 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3578 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3637

Qy        3481 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3638 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3697

Qy        3541 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3698 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3757

Qy        3601 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3758 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3817

Qy        3661 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3818 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3877

Qy        3721 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3878 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3937

Qy        3781 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3938 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3997

Qy        3841 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 3900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3998 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 4057

Qy        3901 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 3960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4058 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 4117

Qy        3961 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4118 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4177

Qy        4021 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4178 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4237

Qy        4081 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4238 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4297

Qy        4141 ACGCCATCACCATCACAACCAAAAAAAAATGT 4172
               ||||||||||||||||||||||||||||||||
Db        4298 ACGCCATCACCATCACAACCAAAAAAAAATGT 4329
```

```
RESULT 5
ABV28244
ID   ABV28244 standard; cDNA; 7473 BP.
XX
AC   ABV28244;
XX
DT   16-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 28235.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
```

```
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer  -
XX
PS   Claim 1; Page 5869-5870; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ   Sequence 7473 BP; 2458 A; 1244 C; 1479 G; 2285 T; 7 other;

  Query Match             99.8%;  Score 4164;  DB 23;  Length 7473;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 4167;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        158 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 217

Qy         61 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        218 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 277

Qy        121 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        278 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 337

Qy        181 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGGTAAAGATGTTCGCTTACTGGTA 240
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        338 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGATAAAGATGTTCGCTTACTGGTA 397

Qy        241 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        398 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 457

Qy        301 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        458 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 517

Qy        361 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 420
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db        518 AAGAGCCCACAATTCAATAGCTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 577

Qy        421 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        578 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 637

Qy        481 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        638 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 697

Qy        541 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        698 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 757

Qy        601 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        758 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 817

Qy        661 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCACTTTTTTAATCAGGTT 720
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        818 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTTTTAATCAGGTT 877

Qy        721 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        878 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 937

Qy        781 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        938 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 997

Qy        841 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        998 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 1057

Qy        901 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1058 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 1117

Qy        961 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1118 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1177

Qy       1021 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1178 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1237

Qy       1081 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1238 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1297

Qy       1141 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1298 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1357

Qy       1201 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1358 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1417

Qy       1261 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1418 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1477

Qy       1321 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1478 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1537

Qy       1381 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1538 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1597

Qy       1441 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1598 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1657

Qy       1501 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1658 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1717

Qy       1561 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1718 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1777

Qy       1621 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1778 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1837

Qy       1681 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1838 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1897

Qy       1741 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1898 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1957

Qy   1801 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1958 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 2017

Qy   1861 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2018 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 2077

Qy   1921 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2078 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 2137

Qy   1981 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2138 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2197

Qy   2041 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2198 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2257

Qy   2101 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2258 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2317

Qy   2161 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2318 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2377

Qy   2221 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2378 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2437

Qy   2281 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2438 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2497

Qy   2341 TTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2400
          ||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||
Db   2498 TTTGCTGCTCCTTTGAAATCTTTGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2557

Qy   2401 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2558 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2617

Qy   2461 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2618 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2677

Qy   2521 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2678 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2737

Qy   2581 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2738 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2797

Qy   2641 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2798 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2857

Qy   2701 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2858 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2917

Qy   2761 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2918 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2977

Qy   2821 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2978 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 3037

Qy   2881 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3038 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 3097

Qy   2941 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3098 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3157

Qy   3001 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3158 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3217

Qy   3061 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3218 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3277

Qy   3121 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3278 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3337

Qy   3181 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3338 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3397

Qy   3241 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3398 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3457

Qy   3301 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3458 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3517

Qy   3361 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3518 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3577

Qy   3421 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3578 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3637

Qy   3481 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3638 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3697

Qy   3541 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3698 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3757

Qy   3601 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3758 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3817

Qy   3661 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3818 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3877

Qy   3721 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3878 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3937

Qy   3781 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3938 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3997

Qy   3841 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3998 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 4057

Qy   3901 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4058 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 4117
```

```
Qy     3961 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4118 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4177

Qy     4021 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4178 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4237

Qy     4081 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4238 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4297

Qy     4141 ACGCCATCACCATCACAACCAAAAAAAAATGT 4172
            ||||||||||||||||||||||||||||||||
Db     4298 ACGCCATCACCATCACAACCAAAAAAAAATGT 4329


RESULT 6
ABX14056
ID   ABX14056 standard; cDNA; 6744 BP.
XX
AC   ABX14056;
XX
DT   24-FEB-2003  (first entry)
XX
DE   cDNA encoding human SCC-112 tumour supressor gene.
XX
KW   ss; gene; human; chromosome 4p14; SCC-112; cancer; apoptosis; allodynia;
KW   degenerative disorder; metastasis inhibition; breast cancer; causalgia;
KW   kidney cancer; bladder cancer; pancreatic cancer; colon cancer;
KW   squamous cell carcinoma; head trauma; spinal cord injury; herpes zoster;
KW   global and focal ischaemic and haemorrhagic stroke; epilepsy; neuralgia;
KW   hypoxia-induced nerve cell damage; anxiety; diabetes mellitus;
KW   cardiac arrest; spinal cord lesion; stomach cancer; lung cancer;
KW   neonatal distress; Alzheimer's disease; uterine cancer; colon cancer;
KW   multiple sclerosis; phantom limb pain; hyperalgesia; Down's syndrome;
KW   Huntington's disease; Parkinson's disease; Korsakoff's syndrome;
KW   amyotrophic lateral sclerosis; cell survival; cell proliferation;
KW   tumour supressor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             232..4125
FT                   /*tag=  a
FT                   /product= "Human SCC-112"
FT   polyA_signal    6721..6726
FT                   /*tag=  b
FT                   /standard_name= "PolyA signal"
XX
PN   WO200281641-A2.
XX
PD   17-OCT-2002.
XX
PF   08-APR-2002; 2002WO-US10850.
XX
PR   06-APR-2001; 2001US-281780P.
XX
PA   (GEOU ) UNIV GEORGETOWN.
XX
PI   Kasid UN,  Kumar D,  Ahmad I;
XX
DR   WPI; 2003-103330/09.
DR   P-PSDB; ABG72803.
XX
PT   New isolated SSC (undefined) tumor suppressor polypeptides and
PT   polynucleotides, useful for diagnosing, preventing or treating cancer
PT   or degenerative disease, e.g. Alzheimer's Disease, Huntington's
PT   disease, or multiple sclerosis  -
XX
PS   Claim 1; Fig 1B; 83pp; English.
XX
CC   The invention relates to a new isolated polypeptide SCC-112. The SCC-112
CC   polypeptides and polynucleotides are useful for diagnosing, preventing or
CC   treating cancer (e.g. breast cancer, kidney cancer, bladder cancer,
CC   pancreatic cancer, colon cancer, squamous cell carcinoma, uterine
CC   cancer, stomach cancer, colon cancer, lung cancer); or degenerative
CC   disease or disorder (e.g. global and focal ischaemic and haemorrhagic
CC   stroke, head trauma, spinal cord injury, hypoxia-induced nerve cell
CC   damage, nerve cell damage caused by cardiac arrest or neonatal distress,
CC   epilepsy, anxiety, diabetes mellitus, multiple sclerosis, phantom limb
CC   pain, causalgia, neuralgias, herpes zoster, spinal cord lesions,
CC   hyperalgesia, allodynia, Alzheimer's disease, Huntington's disease,
CC   Parkinson's disease, amyotrophic lateral sclerosis, Down's syndrome and
CC   Korsakoff's syndrome). The polypeptides and polynucleotides are also
CC   useful for inducing apoptosis in cancer cells, increasing survival or
CC   proliferation of a cell, or inhibiting cancer cell proliferation and/or
CC   metastasis in a cancer patient. The polynucleotides can be used as probes
CC   to detect complementary nucleotide sequences, or as primers to obtain
CC   additional copies of the polynucleotides. SSC-112 may also be used for
CC   identifying drugs for treatment of cancers. The present sequence
CC   represents cDNA encoding the human SCC-112 tumour supressor gene which
CC   is located on chromosome 4p14.
XX
SQ   Sequence 6744 BP; 2156 A; 1243 C; 1400 G; 1945 T; 0 other;

  Query Match             43.5%;  Score 1815.6;  DB 25;  Length 6744;
  Best Local Similarity   69.7%;  Pred. No. 0;
  Matches 2457;  Conservative    0;  Mismatches 1069;  Indels    0;  Gaps    0;

Qy       32 GAAAAATTACATATCCGCCTGGGGTCAAGGAAATATCAGATAAAATATCTAAAGAGGAGA 91
            | || ||  | || || || ||||| || || ||  | || || ||  | |  || ||||
Db      173 GGAAGATCGCTTACCCTCCGGGGGTAAAAGAGATCACCGACAAGATCACCACGGACGAGA 232

Qy       92 TGGTGAGACGATTAAAGATGGTTGTGAAAACTTTTATGGATATGGACCAGGACTCTGAAG 151
            || | | |||  | ||||||||| ||||||||| |||||||||||||||| |||||||| ||||
Db      233 TGATCAAACGCCTGAAGATGGTAGTGAAAACCTTTATGGATATGGATCAGGACTCAGAAG 292

Qy      152 AAGAAAAGGAGCTTTATTTAAACCTAGCTTTACATCTTGCTTCAGATTTTTTTCTCAAGC 211
            | |||||  |||  ||| |    ||||| || |||||||| || || || || |||| | 
Db      293 ATGAAAAACAGCAGTATCTCCCACTAGCCTTGCATCTTGCATCTGAATTCTTCCTCAGGA 352

Qy      212 ATCCTGGTAAAGATGTTCGCTTACTGGTAGCCTGCTGCCTTGCTGATATTTTCAGGATTT 271
            | ||   |||||||| ||  | || |||||| || ||  | |||||||| ||  | || |
Db      353 ACCCCAATAAAGATGTGCGTCTCCTTGTAGCATGTTGTTTGGCTGATATCTTTCGTATCT 412

Qy      272 ATGCTCCTGAAGCTCCTTACACATCCCCTGATAAACTAAAGGATATATTTATGTTTATAA 331
            |||| || |||||||| || || |||| ||||||||| ||||| |||||| ||||||| |
Db      413 ATGCCCCAGAAGCTCCATATACTTCCCATGATAAACTTAAGGACATATTTTTGTTTATTA 472

Qy      332 CAAGACAGTTGAAGGGGCTAGAGGATACAAAGAGCCCACAATTCAATAGGTATTTTTATT 391
            | ||||| || || ||  | |||||||||||||| ||||| || ||||| || |||||||
Db      473 CCAGACAATTAAAAGGTTTGGAGGATACAAAGAGTCCACAGTTTAATAGATACTTTTATT 532

Qy      392 TACTTGAGAACATTGCTTGGGTCAAGTCATATAACATATGCTTTGAGTTAGAAGATAGCA 451
            || | |||||  | |||||||| || |||||||||||| ||||||||| || |||||| |||
Db      533 TATTAGAGAATTTAGCTTGGGTTAAATCATATAACATCTGCTTTGAATTGGAAGATTGCA 592

Qy      452 ATGAAATTTTCACCCAGCTATACAGAACCTTATTTTCAGTTATAAACAATGGCCACAATC 511
            |||||||||| |  ||||| |  |||||  | || ||||| || |||||| |||||||| 
Db      593 ATGAAATTTTTATTCAGCTTTTTAGAACTCTCTTCTCAGTGATCAACAATAGCCACAATA 652

Qy      512 AGAAAGTCCATATGCACATGGTAGACCTTATGAGCTCTATTATTTGTGAAGGTGATACAG 571
            |||| || || ||||||||| ||||  | ||||| ||||| ||    |||||||||  ||
Db      653 AGAAGGTACAAATGCACATGCTAGATTTGATGAGTTCTATCATCATGGAAGGTGATGGAG 712

Qy      572 TGTCTCAGGAGCTTTTGGATACGGTTTTAGTAAATCTGGTACCTGCTCATAAGAATTTAA 631
            |  |||| ||  | ||||   |  || |  |||      | ||||| |||||||| ||||
Db      713 TTACTCAAGAATTATTGGGCTCCATTCTTATTAACCTCATTCCTGCACATAAGAACTTAA 772

Qy      632 ACAAGCAAGCATATGATTTGGCAAAGGCTTTACTGAAGAGGACAGCTCAAGCTATTGAGC 691
            | || ||  | | |||  | ||||| |   || |||| || ||||  ||  |||||||| 
Db      773 ATAAACAGTCCTTTGACCTTGCAAAAGTGCTATTGAAAAGAACAGTCCAGACTATTGAGG 832

Qy      692 CATATATTACCACTTTTTTAATCAGGTTCTGATGCTTGGGAAAACATCTATCAGCGATT 751
            |||  ||| |||| |||||| ||||| || ||| |||| || | | ||||  | || ||||
```

```
Db    833 CATGCATTGCCAATTTTTTCAATCAAGTCCTGGTGCTGGGAAGATCATCAGTAAGTGATT 892

Qy    752 TGTCAGAGCATGTCTTTGACTTAATTTTGGAGCTCTACAATATTGATAGTCATTTGCTGC 811
Db    893 TGTCAGAACATGTATTTGATCTGATTCAGGAACTTTTTGCTATAGATCCTCATTTATTAT 952

Qy    812 TCTCTGTTTTACCCCAGCTTGAATTTAAATTAAAGAGCAATGATAATGAGGAGCGCCTAC 871
Db    953 TATCCGTCATGCCACAGCTTGAATTCAAACTAAAGAGCAATGATGGAGAAGAGCGATTAG 1012

Qy    872 AAGTTGTTAAACTACTGGCAAAAATGTTTGGGGCAAAGGATTCAGAATTGGCTTCTCAAA 931
Db   1013 CTGTTGTTCGACTTCTAGCTAAATTGTTTGGCTCCAAAGATTCTGATTTGGCAACACAGA 1072

Qy    932 ACAAGCCACTTTGGCAGTGCTACTTGGGCAGGTTTAATGATATCCATGTACCAATCCGCC 991
Db   1073 ATCGTCCTCTTTGGCAATGTTTTCTTGGACGATTTAATGATATTCATGTTCCTGTGAGAT 1132

Qy    992 TGGAATGTGTGAAATTTGCTAGCCATTGTCTCATGAACCATCCTGATTTAGCAAAAGACT 1051
Db   1133 TAGAAAGTGTGAAATTTGCCAGTCATTGTTTAATGAATCACCCAGATTTAGCGAAGGATC 1192

Qy   1052 TAACAGAGTATCTTAAAGTGAGGTCACATGACCCTGAGGAAGCTATTAGACATGATGTTA 1111
Db   1193 TCACAGAATATTTAAAGGTTAGATCACATGATCCAGAAGAAGCTATTCGTCATGATGTCA 1252

Qy   1112 TTGTGTCAATAGTTACAGCTGCTAAAAAGGATATTCTTCTGGTCAATGATCACTTACTTA 1171
Db   1253 TTGTTACTATAATAACAGCTGCCAAGAGGGACCTGGCCTTAGTAAATGATCAGCTGCTTG 1312

Qy   1172 ATTTTGTGAGAGAGAGAACATTAGACAAACGATGGAGAGTACGCAAAGAAGCCATGATGG 1231
Db   1313 GCTTTGTAAGGGAAAGAACACTGGATAAACGGTGGCGAGTAAGAAAAGAAGCTATGATGG 1372

Qy   1232 GACTTGCCCAAATTTATAAGAAATATGCTTTACAGTCAGCAGCTGGAAAAGATGCTGCAA 1291
Db   1373 GTCTGGCTCAGCTTTATAAGAAATACTGTCTTCATGGTGAAGCAGGAAAGGAAGCTGCAG 1432

Qy   1292 AACAGATAGCATGGATCAAAGACAAATTGCTACATATATATTATCAAAATAGTATTGATG 1351
Db   1433 AGAAAGTCAGCTGGATAAAGGACAAACTTCTGCATATTTATTATCAGAACAGCATTGACG 1492

Qy   1352 ATCGACTACTTGTTGAACGGATCTTTGCTCAATACATGGTTCCTCACAATTTAGAAACTA 1411
Db   1493 ACAAACTGTTGGTAGAGAAAATCTTTGCTCAGTATCTTGTCCCCCACAACCTGGAAACAG 1552

Qy   1412 CAGAACGGATGAAATGCTTATATTACTTGTATGCCACACTGGATTTAAATGCTGTGAAAG 1471
Db   1553 AAGAGAGAATGAAATGCTTATATTACTTATATGCTAGTTTGGATCCAAATGCTGTAAAAG 1612

Qy   1472 CATTGAATGAAATGTGGAAATGTCAAAATCTGCTCCGACATCAAGTAAAGGATTTGCTTG 1531
Db   1613 CTCTCAACGAAATGTGGAAGTGTCAGAACATGCTTCGGAGCCATGTACGCGAACTATTGG 1672

Qy   1532 ACTTGATTAAGCAACCCAAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGG 1591
Db   1673 ATTTGCACAAGCAGCCTACATCAGAGGCTAACTGTTCTGCCATGTTTGGAAAACTGATGA 1732

Qy   1592 TTATTACAAGAAATTTACCTGATCCTGGTAAGGCTCAGGATTTCATGAAGAAATTCACAC 1651
Db   1733 CCATAGCAAAGAATTTGCCTGACCCCGGGAAAGCACAAGATTTTGTGAAGAAATTTAACC 1792

Qy   1652 AGGTGTTAGAAGATGATGAGAAAATAAGAAAGCAGTTAGAAGTACTTGTTAGTCCAACAT 1711
Db   1793 AGGTTCTCGGCGATGATGAGAAACTTCGGTCTCAGTTGGAGTTATTAATTAGCCCAACCT 1852

Qy   1712 GCTCCTGCAAGCAGGCTGAAGGTTGTGTGCGTGAAATAACTAAGAAGTTGGGCAACCCCA 1771
Db   1853 GTTCTTGCAAACAAGCAGATATTTGTGTGAGAGAAATAGCCCGGAAACTTGCAAATCCTA 1912

Qy   1772 AACAGCCTACAAATCCTTTCCTGGAAATGATCAAGTTTCTCTTGGAGAGGATAGCACCTG 1831
Db   1913 AGCAACCAACAAATCCTTTTCTAGAGATGGTCAAATTTCTGTTGGAAAGAATCGCACCTG 1972

Qy   1832 TGCACATAGATACCGAATCTATCAGTGCTCTTATTAAACAAGTGAACAAATCAATAGATG 1891
Db   1973 TGCACATTGATTCAGAAGCCATAAGTGCACTAGTGAAATTGATGAATAAGTCAATAGAGG 2032

Qy   1892 GAACAGCAGATGATGAAGATGAGGGTGTTCCAACTGATCAAGCCATCAGAGCAGGTCTTG 1951
Db   2033 GGACAGCAGATGATGAAGAGGAGGGTGTAAGTCCAGATACAGCTATCCGTTCAGGACTTG 2092

Qy   1952 AACTGCTTAAGGTACTCTCATTTACACATCCCATCTCATTTCATTCTGCTGAAACATTTG 2011
Db   2093 AACTTCTTAAGGTTCTGTCTTTTACACATCCTACCTCGTTCCACTCTGCAGAGACATATG 2152

Qy   2012 AATCATTACTGGCTTGTCTGAAAATGGATGATGAAAAAGTAGCAGAAGCTGCACTACAAA 2071
Db   2153 AGTCCTTGTTACAGTGCCTAAGAATGGAGGATGACAAGGTAGCAGAAGCTGCTATTCAAA 2212

Qy   2072 TTTTCAAAAACACAGGAAGCAAAATTGAAGAGGATTTCCACACATCAGATCAGCCTTGC 2131
Db   2213 TTTTTAGAAATACAGGTCACAAAATAGAAACAGACCTTCCCCAGATACGATCGACCTTAA 2272

Qy   2132 TTCCTGTTTTACATCACAAATCTAAAAAAGGACCCCCCCGTCAAGCCAAATATGCCATTC 2191
Db   2273 TTCCCATTTTACATCAAAAAGCAAAGAGGGGTACTCCACACCAAGCAAAACAGGCTGTGC 2332

Qy   2192 ATTGTATCCATGCGATATTTTCTAGTAAAGAGACCCAGTTTGCACAGATATTTGAGCCTC 2251
Db   2333 ACTGTATACACGCCATATTCACAAATAAAGAAGTCCAGCTTGCACAGATTTTTGAGCCAC 2392

Qy   2252 TGCATAAGAGCCTAGATCCAAGCAACCTGGAACATCTCATAACACCATTGGTTACTATTG 2311
Db   2393 TCAGTAGGAGTCTGAATGCTGATGTGCCAGAACAACTTATAACTCCATTAGTTTCATTGG 2452

Qy   2312 GTCATATTGCTCTCCTTGCACCTGATCAATTTGCTGCTCCTTGGAAATCTTGGGTAGCTA 2371
Db   2453 GCCACATTTCTATGTTAGCACCAGATCAGTTTGCTTCCCCAATGAAATCTGTAGTAGCAA 2512

Qy   2372 CTTTCATTGTGAAAGATCTTCTCATGAATGATCGGCTTCCAGGGAAAAAGACAACTAAAC 2431
Db   2513 ATTTTATTGTGAAAGATCTGCTAATGAATGACAGGTCAACAGGTGAAAAGAATGGAAAAC 2572

Qy   2432 TTTGGGTTCCAGATGAAGAAGTATCTCCTGAGACAATGGTCAAAATTCAGGCTATTAAAA 2491
Db   2573 TGTGGTCTCCAGATGAAGAGGTTTCCCCTGAAGTACTAGCAAAGGTACAGGCAATTAAAC 2632

Qy   2492 TGATGGTTCGATGGCTACTTGGAATGAAAAATAATCACAGTAAATCAGGAACTTCTACCT 2551
Db   2633 TTCTGGTAAGGTGGCTGTTGGGTATGAAAAACAACCAGTCTAAATCTGCCAATTCAACCC 2692

Qy   2552 TAAGATTGCTAACAACAATATTGCATAGTGATGGAGACTTGACAGAACAGGGGAAAATTA 2611
Db   2693 TTCGGTTATTATCAGCGATGTTGGTTAGTGAGGGTGACCTGACAGAGCAAAAGAGGATCA 2752

Qy   2612 GTAAACCAGATATGTCACGTCTGAGACTTGCTGCTGGGAGTGCTATTGTGAAGCTGGCAC 2671
Db   2753 GTAAATCTGATATGTCTCGCTTGCGATTAGCTGCTGGTAGTGCCATAATGAAGCTTGCTC 2812

Qy   2672 AAGAACCCTGTTACCATGAAATCATCACATTAGAACAATATCAGCTATGTGCATTAGCTA 2731
Db   2813 AGGAACCTTGTTACCATGAAATTATTACCCCAGAACAGTTTCAGCTCTGTGCACTTGTTA 2872

Qy   2732 TCAACGATGAATGCTATCAAGTAAGACAAGTGTTTGCCCAGAAACTTCACAAAGGCCTTT 2791
Db   2873 TTAATGATGAGTGTTACCAAGTAAGGCAGATATTTGCTCAGAAGCTGCATAAGGCACTTG 2932

Qy   2792 CCCGTTTACGGCTTCCACTTGAGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTG 2851
Db   2933 TGAAGTTACTGCTCCCATTGGAGTATATGGCGATCTTTGCCTTGTGTGCCAAAGATCCTG 2992

Qy   2852 TAAAGGAGAGAAGAGCTCATGCTAGGCAATGTTTGGTGAAAAATATAAATGTAAGGCGGG 2911
Db   2993 TGAAGGAGAGAAGAGCACACGCACGACAATGTTTACTGAAAAATATCAGTATACGCAGGG 3052
```

```
Qy       2912 AGTATCTGAAGCAGCATGCAGCTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATG 2971
Db       3053 AATACATTAAGCAGAATCCTATGGCTACTGAGAAATTATTATCACTGTTGCCTGAATATG 3112

Qy       2972 TTGTTCCATATACAATTCACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATA 3031
Db       3113 TAGTTCCATACATGATTCACCTGCTAGCCCATGATCCAGATTTTACAAGATCACAAGATG 3172

Qy       3032 TTGAACAACTTAAAGATGTTAAAGAATGTCTTTGGTTTGTTCTGGAAATATTAATGGCTA 3091
Db       3173 TTGATCAGCTTCGTGATATCAAAGAGTGCCTATGGTTCATGCTTGAAGTTTTAATGACAA 3232

Qy       3092 AAAATGAAAATAACAGTCACGCTTTTATCAGAAAGATGGTAGAAAATATTAAACAAACAA 3151
Db       3233 AGAATGAAAACAATAGCCATGCCTTTATGAAGAAGATGGCAGAGAACATCAAGTTAACCA 3292

Qy       3152 AAGATGCCCAAGGACCAGATGATGCAAAAATGAATGAAAAACTGTACACTGTGTGTGATG 3211
Db       3293 GAGATGCCCAGTCTCCAGATGAATCCAAGACAAATGAAAAACTGTATACAGTATGTGATG 3352

Qy       3212 TTGCCATGAATATCATCATGTCAAAGAGTACTACATACAGTTTGGAATCTCCTAAAGACC 3271
Db       3353 TGGCTCTCTGTGTTATAAATAGTAAAAGTGCTTTGTGCAATGCAGATTCACCAAAGGACC 3412

Qy       3272 CGGTACTACCAGCTCGTTTCTTCACTCAACCTGACAAGAATTTCAGTAACACCAAAAATT 3331
Db       3413 CAGTCCTCCCAATGAAATTTTTTACACAACCTGAAAAGGACTTCTGTAACGATAAGAGTT 3472

Qy       3332 ATCTGCCTCCTGAAATGAAATCATTTTTCACTCCTGGAAAACCTAAAACAACCAATGTTC 3391
Db       3473 ATATTTCAGAAGAGACAAGAGTACTTCTGTTAACAGGAAAGCCAAAGCCTGCTGGAGTAC 3532

Qy       3392 TAGGAGCTGTTAACAAGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAAATCATCAC 3451
Db       3533 TAGGTGCAGTAAATAAGCCTTTATCAGCAACGGGAAGGAAACCCTATGTTAGAAGCACTG 3592

Qy       3452 GAATGGAAACTGTAAGCAATGCAAGCAGCAGCTCAAATCCAAGCTCTCCTGGAAGAATAA 3511
Db       3593 GCACTGAGACTGGAAGCAATATTAATGTAAATTCAGAGCTGAACCCTTCAACCGGAAATC 3652

Qy       3512 AGGGGAGGCTTGATAGTTCTGAAATGGATCACAGTGAAAATGAAGA 3557
Db       3653 GATCAAGGGAACAGAGTTCAGAGGCAGCAGAAACTGGAGTTAGTGA 3698
```

RESULT 7
ABQ99271
ID   ABQ99271 standard; cDNA; 5469 BP.
XX
AC   ABQ99271;
XX
DT   25-FEB-2003 (first entry)
XX
DE   Human coding sequence SEQ ID 4.
XX
KW   Human; expressed sequence tag; EST; chromosome 4;
KW   haematopoietic disorder; central nervous system disease; viral infection;
KW   peripheral nervous system disease; non-healing wound; infectious disease;
KW   immune deficiency; immune disorder; bacterial infection; allergy; cancer;
KW   fungal infection; autoimmune disorder; coagulation disorder; nootropic;
KW   antiallergic; antiinflammatory; immunosuppressive; neuroprotective;
KW   cytostatic; haemostatic; virucide; antibacterial; fungicide;
KW   immunostimulant; cerebroprotective; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200259260-A2.
XX
PD   01-AUG-2002.
XX
PF   16-NOV-2001; 2001WO-US42950.
XX
PR   17-NOV-2000; 2000US-0714936.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;
PI   Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR   WPI; 2002-590824/63.
DR   N-PSDB; ABP64685.
XX
PT   New isolated polynucleotide, useful in research, diagnostic or
PT   therapeutic methods, e.g. preventing or treating disorders involving
PT   aberrant protein expression or biological activity -
XX
PS   Claim 1; SEQ ID 4; 394pp; English.
XX
CC   The present invention relates to novel human coding sequences
CC   (ABQ99268-ABQ99608) and proteins (ABP64682-ABP65022). The sequences are
CC   useful in therapeutic, diagnostic and research methods. The
CC   polynucleotides may be used in the field of molecular biology as
CC   hybridisation probes, primers for PCR, for chromosome and gene mapping,
CC   for the recombinant production of protein, or in generation of anti-sense
CC   DNA or RNA. The polynucleotides are useful in diagnostics as expressed
CC   sequence tags (ESTs) for identifying expressed genes or for physical
CC   mapping of the human genome. The proteins may be used as molecular weight
CC   markers, or as nutritional sources or supplements. The proteins may be
CC   used to maintain and expand cell population in a totipotential or
CC   pluripotential state useful for re-engineering damaged or diseased
CC   tissues, transplantation, manufacture of bio-pharmaceuticals or the
CC   development of bio-sensors. The polynucleotides and proteins are useful
CC   for preventing, treating or ameliorating disorders involving aberrant
CC   protein expression or biological activity, e.g. haematopoietic disorders,
CC   central/peripheral nervous system diseases, mechanical and traumatic
CC   disorders, non-healing wounds, immune deficiencies and disorders,
CC   infectious diseases caused by viral, bacterial or fungal infection,
CC   autoimmune disorders, allergic reactions and conditions, coagulation
CC   disorders, or cancer. The polynucleotide sequences of the invention were
CC   assembled from ESTs isolated mainly by sequencing by hybridisation, and
CC   in some cases, sequences obtained from one or more public databases.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 5469 BP; 1801 A; 983 C; 1097 G; 1588 T; 0 other;

Query Match 27.7%; Score 1157.2; DB 24; Length 5469;
Best Local Similarity 68.2%; Pred. No. 4.4e-275;
Matches 1606; Conservative 0; Mismatches 748; Indels 0; Gaps 0;

```
Qy       1204 TGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCTTTA 1263
Db         54 TGGCGAGTAAGAAAAGAAGCTATGATGGGTCTGGCTCAGCTTTATAAGAAATACTGTCTT 113

Qy       1264 CAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTGCTA 1323
Db        114 CATGGTGAAGCAGGAAAGGAAGCTGCAGAGAAAGTCAGCTGGATAAAGGACAAACTTCTG 173

Qy       1324 CATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCTCAA 1383
Db        174 CATATTTATTATCAGAACAGCATTGACGACAAACTGTTGGTAGAGAAAATCTTTGCTCAG 233

Qy       1384 TACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTGTAT 1443
Db        234 TATCTTGTCCCCCACAACCTGGAAACAGAAGAGAGAATGAAATGCTTATATTACTTATAT 293

Qy       1444 GCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAATCTG 1503
Db        294 GCTAGTTTGGATCCAAATGCTGTAAAAGCTCTCAACGAAATGTGGAAGTGTCAGAACATG 353

Qy       1504 CTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCCAGT 1563
Db        354 CTTCGGAGCCATGTACGCGAACTATTGGATTTGCACAAGCAGCCTACATCAGAGGCTAAC 413

Qy       1564 GTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGTAAG 1623
```

```
Db     414 TGTTCTGCCATGTTTGGAAAACTGATGACCATAGCAAAGAATTTGCCTGACCCCGGGAAA 473

Qy    1624 GCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGAAAG 1683
Db     474 GCACAAGATTTTGTGAAGAAATTTAACCAGGTTCTCGGCGATGATGAGAAACTTCGGTCT 533

Qy    1684 CAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTGCGT 1743
Db     534 CAGTTGGAGTTATTAATTAGCCCAACCTGTTCTTGCAAACAAGCAGATATTTGTGTGAGA 593

Qy    1744 GAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATGATC 1803
Db     594 GAAATAGCCCGGAAACTTGCAAATCCTAAGCAACCAACAAATCCTTTTCTAGAGATGGTC 653

Qy    1804 AAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCTCTT 1863
Db     654 AAATTTCTGTTGGAAAGAATCGCACCTGTGCACATTGATTCAGAAGCCATAAGTGCGCTA 713

Qy    1864 ATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTTCCA 1923
Db     714 GTGAAATTGATGAATAAGTCAATAGAGGGGACAGCAGATGATGAAGAGGAGGGTGTAAGT 773

Qy    1924 ACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACATCCC 1983
Db     774 CCAGATACAGCTATCCGTTCAGGACTTGAACTTCTTAAGGTTCTGTCTTTTACACATCCT 833

Qy    1984 ATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGATGAT 2043
Db     834 ACCTCGTTCCACTCTGCAGAGACATATGAGTCCTTGTTACAGTGCCTAAGAATGGAGGAT 893

Qy    2044 GAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAAGAG 2103
Db     894 GACAAGGTAGCAGAAGCTGCTATTCAAATTTTTAGAAATACAGGTCACAAAATAGAAACA 953

Qy    2104 GATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAAGGA 2163
Db     954 GACCTTCCCCAGATACGATCGACCTTAATTCCCATTTTACATCAAAAAGCAAAGAGGGGT 1013

Qy    2164 CCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTCTAGTAAAGAG 2223
Db    1014 ACTCCACACCAAGCAAAACAGGCTGTGCACTGTATACACGCCATATTCACAAATAAAGAA 1073

Qy    2224 ACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTGGAA 2283
Db    1074 GTCCAGCTTGCACAGATTTTTGAGCCACTCAGTAGGAGTCTGAATGCTGATGTGCCAGAA 1133

Qy    2284 CATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAATTT 2343
Db    1134 CAACTTATAACTCCATTAGTTTCATTGGGCCACATTTCTATGTTAGCACCAGATCAGTTT 1193

Qy    2344 GCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAATGAT 2403
Db    1194 GCTTCCCCAATGAAATCTGTAGTAGCAAATTTTATTGTGAAAGATCTGCTAATGAATGAC 1253

Qy    2404 CGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCTGAG 2463
Db    1254 AGGTCAACAGGTGAAAAGAATGGAAAACTGTGGTCTCCAGATGAAGAGGTTTCCCCTGAA 1313

Qy    2464 ACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAAAAT 2523
Db    1314 GTACTAGCAAAGGTACAGGCAATTAAACTTCTGGTAAGGTGGCTGTTGGGTATGAAAAAC 1373

Qy    2524 AATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGTGAT 2583
Db    1374 AACCAGTCTAAATCTGCCAATTCAACCCTTCGGTTATTATCAGCGATGTTGGTTAGTGAG 1433

Qy    2584 GGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTTGCT 2643
Db    1434 GGTGACCTGACAGAGCAAAAGAGGATCAGTAAATCTGATATGTCTCGCTTGCGATTAGCT 1493

Qy    2644 GCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACATTA 2703
Db    1494 GCTGGTAGTGCCATAATGAAGCTTGCTCAGGAACCTTGTTACCATGAAATTATTACCCCA 1553

Qy    2704 GAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAAGTG 2763
Db    1554 GAACAGTTTCAGCTCTGTGCACTTGTTATTAATGATGAGTGTTACCAAGTAAGGCAGATA 1613

Qy    2764 TTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATGGCA 2823
Db    1614 TTTGCTCAGAAGCTGCATAAGGCACTTGTGAAGTTACTGCTCCCATTGGAGTATATGGCG 1673

Qy    2824 ATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAATGT 2883
Db    1674 ATCTTTGCCTTGTGTGCCAAAGATCCTGTGAAGGAGAGAAGAGCACACGCACGACAATGT 1733

Qy    2884 TTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGTGAA 2943
Db    1734 TTACTGAAAAATATCAGTATACGCAGGGAATACATTAAGCAGAATCCTATGGCTACTGAG 1793

Qy    2944 AAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCACAT 3003
Db    1794 AAATTATTATCACTGTTGCCTGAATATGTAGTTCCATACATGATTCACCTGCTAGCCCAT 1853

Qy    3004 GACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGTCTT 3063
Db    1854 GATCCAGATTTTACAAGATCACAAGATGTTGATCAGCTTCGTGATATCAAAGAGTGCCTA 1913

Qy    3064 TGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATCAGA 3123
Db    1914 TGGTTCATGCTTGAAGTTTTAATGACAAAGAATGAAAACAATAGCCATGCCTTTATGAAG 1973

Qy    3124 AAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAAATG 3183
Db    1974 AAGATGGCAGAGAACATCAAGTTAACCAGAGATGCCCAGTCTCCAGATGAATCCAAGACA 2033

Qy    3184 AATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGTACT 3243
Db    2034 AATGAAAAACTGTATACAGTATGTGATGTGGCTCTCTGTGTTATAAATAGTAAAAGTGCT 2093

Qy    3244 ACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAACCT 3303
Db    2094 TTGTGCAATGCAGATTCACCAAAGGACCCAGTCCTCCCAATGAAATTTTTTACACAACCT 2153

Qy    3304 GACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTCACT 3363
Db    2154 GAAAAGGACTTCTGTAACGATAAGAGTTATATTTCAGAAGAGACAAGAGTACTTCTGTTA 2213

Qy    3364 CCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCAGCA 3423
Db    2214 ACAGGAAAGCCAAAGCCTGCTGGAGTACTAGGTGCAGTAAATAAGCCTTTATCAGCAACG 2273

Qy    3424 GGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGCAGC 3483
Db    2274 GGAAGGAAACCCTATGTTAGAAGCACTGGCACTGAGACTGGAAGCAATATTAATGTAAAT 2333

Qy    3484 TCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGATCAC 3543
Db    2334 TCAGAGCTGAACCCTTCAACCGGAAATCGATCAAGGGAACAGAGTTCAGAGGCAGCAGAA 2393

Qy    3544 AGTGAAAATGAAGA 3557
Db    2394 ACTGGAGTTAGTGA 2407
```

RESULT 8
AAA47423
ID AAA47423 standard; DNA; 3957 BP.
XX
AC AAA47423;
XX
DT 20-OCT-2000 (first entry)
XX
DE Sequence encoding human neuron-associated protein.
XX

```
KW   Neuron associated protein; NEUAP; neurological disorder; epilepsy;
KW   ischemic cerebrovascular disease; stroke; cerebral neoplasm;
KW   Alzheimer's disease; Pick's disease; Huntington's disease;
KW   dementia; Parkinson's disease; demyelinating disease; meningitis;
KW   prion disease; kuru; Creutzfeldt-Jakob disease; neurofibromatosis;
KW   cerebral palsy; muscular dystrophy; central nervous system; CNS;
KW   peripheral nervous system; PNS; myopathy; schizophrenia;
KW   actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;
KW   cirrhosis; hepatitis; mixed connective tissue disease; MCTD;
KW   myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;
KW   autoimmune disease; inflammation; acquired immunodeficiency syndrome;
KW   AIDS; Addison's disease; adult respiratory distress syndrome;
KW   allergy; ankylosing spondylitis; amyloidosis; anaemia; asthma;
KW   Werner syndrome, trauma; human; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             140..2704
FT                   /*tag= a
FT                   /product= Neuron associated protein
XX
PN   WO200034477-A2.
XX
PD   15-JUN-2000.
XX
PF   10-DEC-1999;   99WO-US30408.
XX
PR   11-DEC-1998;   98US-0210083.
PR   11-DEC-1998;   98US-9123456.
PR   09-FEB-1999;   99US-0119365.
PR   16-MAR-1999;   99US-0124687.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Tang YT,  Yue H,  Baughn MR,  Hillman JL,  Lal P,  Au-young J,  Yang J;
PI   Lu DAM,  Azimzai Y;
XX
DR   WPI; 2000-423423/36.
DR   P-PSDB; AAB01382.
XX
PT   New human neuron-associated proteins and polynucleotides encoding them,
PT   useful for diagnosis, treatment and prevention of cell proliferative
PT   disorders including cancer, neuronal and neurological disorders
XX
PS   Claim 9; Page 130-131; 145pp; English.
XX
CC   Human neuron-associated proteins (NEUAP) can be used for for
CC   treating or preventing a disorder associated with decreased
CC   expression or activity of NEUAP. Antagonists of NEUAP are useful for
CC   treating or preventing disorder associated with increased expression
CC   or activity of NEUAP. NEUAP or their fragments or derivatives are
CC   useful for treating neurological disorder such as epilepsy, ischemic
CC   cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's
CC   disease, Pick's disease, Huntington's disease, dementia and
CC   Parkinson's disease. NEUAPs are also useful for treating other
CC   demyelinating diseases, bacterial and viral meningitis, prion
CC   diseases including kuru, Creutzfeldt-Jakob disease, nutritional and
CC   metabolic diseases of the nervous system, neurofibromatosis, other
CC   developmental disorders of the central nervous system, cerebral
CC   palsy, neuroskeletal disorders, autonomic nervous system disorders,
CC   cranial nerve disorders, spinal cord diseases, muscular dystrophy and
CC   other neuromuscular disorders, peripheral nervous system disorders,
CC   inherited, metabolic, endocrine, and toxic myopathies, mental
CC   disorders including mood, anxiety and schizophrenic disorders, a cell
CC   proliferative disorder such as actinic keratosis, arteriosclerosis,
CC   atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective
CC   tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal
CC   haemoglobinuria, cancers of the adrenal gland, bladder, bone,
CC   bone marrow, brain, breast, cervix, and an autoimmune/inflammatory
CC   disorder such as acquired immunodeficiency syndrome (AIDS), Addison's
CC   disease, adult respiratory distress syndrome, allergies, ankylosing
CC   spondylitis, amyloidosis, anemia, asthma, Werner syndrome,
CC   complications of cancer, hemodialysis, and extracorporeal circulation,
CC   viral, bacterial, fungal parasitic, protozoal, and helminthic
CC   infections, and trauma. This sequence was given the Incyte ID no.
CC   2888437CB1.
XX
SQ   Sequence 3957 BP; 1303 A; 723 C; 811 G; 1120 T; 0 other;

  Query Match             26.7%;  Score 1112.8;  DB 21;  Length 3957;
  Best Local Similarity   68.1%;  Pred. No. 3.5e-264;
  Matches 1549;  Conservative    0;  Mismatches  727;  Indels    0;  Gaps    0;

Qy       1282 GATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTGCTACATATATATTATCAAAAT 1341
              || |||||| |  |  |    ||||| || |||||| | || ||||| |||||| || 
Db          2 GAAGCTGCAGAGAAAGTCAGCTGGATAAAGGACAAACTTCTGCATATTTATTATCAGAAC 61

Qy       1342 AGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCTCAATACATGGTTCCTCACAAT 1401
              || ||||| ||   |||  | || ||    ||||||||||| ||  | || || ||||| 
Db         62 AGCATTGACGACAAACTGTTGGTAGAGAAAATCTTTGCTCAGTATCTTGTCCCCCACAAC 121

Qy       1402 TTAGAAACTACAGAACGGATGAAATGCTTATATTACTTGTATGCCACACTGGATTTAAAT 1461
               | |||||   |||  | ||||||||||||||||||||| ||||| |   |||||  ||||
Db        122 CTGGAAACAGAAGAGAGAATGAAATGCTTATATTACTTATATGCTAGTTTGGATCCAAAT 181

Qy       1462 GCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAATCTGCTCCGACATCAAGTAAAG 1521
              ||||| ||||||  | || ||||||||| ||||| ||  |||| ||    || |||   
Db        182 GCTGTAAAAGCTCTCAACGAAATGTGGAAGTGTCAGAACATGCTTCGGATCCATGTACGC 241

Qy       1522 GATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCCAGTGTCAAGGCCATATTTTCA 1581
              ||  |  | || |||   ||||| || | | |||| || |       | |||| |||  |
Db        242 GAACTATTGGATTTGCACAAGCAGCCTACATCAGAGGCTAACTGTTCTGCCATGTTTGGA 301

Qy       1582 AAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGTAAGGCTCAGGATTTCATGAAG 1641
              ||| |||||   ||  |||  ||||| ||||| || || || || || |||||  |||||
Db        302 AAACTGATGACCATAGCAAAGAATTTGCCTGACCCCGGGAAAGCACAAGATTTTGTGAAG 361

Qy       1642 AAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGAAAGCAGTTAGAAGTACTTGTT 1701
              ||||| |  |||||  | |  ||||||||||||| |  |    ||||| ||  || |  ||
Db        362 AAATTTAACCAGGTTCTCGGCGATGATGAGAAACTTCGGTCTCAGTTGGAGTTATTAATT 421

Qy       1702 AGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTGCGTGAAATAACTAAGAAGTTG 1761
              || ||||| || || ||||| || || ||   |||||||| |||||| |   |||  | 
Db        422 AGCCCAACCTGTTCTTGCAAACAAGCAGATATTTGTGTGAGAGAAATAGCCCGGAAACTT 481

Qy       1762 GGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATGATCAAGTTTCTCTTGGAGAGG 1821
              |  || || || || || ||||||||||| || || ||| |||| ||||| ||||| || 
Db        482 GCAAATCCTAAGCAACCAACAAATCCTTTTCTAGAGATGGTCAAATTTCTGTTGGAAAGA 541

Qy       1822 ATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCTCTTATTAAACAAGTGAACAAA 1881
              || |||||||||||||| ||| | ||| | || ||||| ||  | |||    |||| || 
Db        542 ATCGCACCTGTGCACATTGATTCAGAAGCCATAAGTGCACTAGTGAAATTGATGAATAAG 601

Qy       1882 TCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTTCCAACTGATCAAGCCATCAGA 1941
              |||||||| || |||||||||||||||||| ||||||||     | |||  ||| ||| | 
Db        602 TCAATAGAGGGGACAGCAGATGATGAAGAGGAGGGTGTAAGTCCAGATACAGCTATCCGT 661

Qy       1942 GCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACATCCCATCTCATTTCATTCTGCT 2001
               |||| |||||||| |||||||| || || |||||||||||| | ||| || || ||||| 
Db        662 TCAGGACTTGAACTTCTTAAGGTTCTGTCTTTTACACATCCTACCTCGTTCCACTCTGCA 721

Qy       2002 GAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGATGATGAAAAAGTAGCAGAAGCT 2061
              || |||| ||| || |   |    || || | |||||| |||||| || ||||||||||||
Db        722 GAGACATATGAGTCCTTGTTACAGTGCCTAAGAATGGAGGATGACAAGGTAGCAGAAGCT 781

Qy       2062 GCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAAGAGGATTTTCCACACATCAGA 2121
              ||  | |||||||| | ||| |||||   |||||| |||   ||  |||| || ||  ||
Db        782 GCTATTCAAATTTTTAGAAATACAGGTCACAAAATAGAAACAGACCTTCCCCAGATACGA 841

Qy       2122 TCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAAGGACCCCCCCGTCAAGCCAAA 2181
              ||  ||||  ||||  |||||||||| ||| | ||| | ||  | || |  ||||| |||
Db        842 TCGACCTTAATTCCCATTTTACATCAAAAAGCAAAGAGGGGTACTCCACACCAAGCAAAA 901

Qy       2182 TATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAAGAGACCCAGTTTGCACAGATA 2241
               | ||   |.|| ||||| || || |||||   | | ||||||    |||| ||||||||||
```

```
Db      902 CAGGCTGTGCACTGTATACACGCCATATTCACAAATAAAGAAGTCCAGCTTGCACAGATT 961

Qy     2242 TTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTGGAACATCTCATAACACCATTG 2301
Db      962 TTTGAGCCACTCAGTAGGAGTCTGAATGCTGATGTGCCAGAACAACTTATAACTCCATTA 1021

Qy     2302 GTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAATTTGCTGCTCCTTGGAAATCT 2361
Db     1022 GTTTCATTGGGCCACATTTCTATGTTAGCACCAGATCAGTTTGCTTCCCCAATGAAATCT 1081

Qy     2362 TGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAATGATCGGCTTCCAGGGAAAAAG 2421
Db     1082 GTAGTAGCAAATTTTATTGTGAAAGATCTGCTAATGAATGACAGGTCAACAGGTGAAAAG 1141

Qy     2422 ACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCTGAGACAATGGTCAAAATTCAG 2481
Db     1142 AATGGAAAACTGTGGTCTCCAGATGAAGAGGTTTCCCCTGAAGTACTAGCAAAGGTACAG 1201

Qy     2482 GCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAAAATAATCACAGTAAATCAGGA 2541
Db     1202 GCAATTAAACTTCTGGTAAGGTGGCTGTTGGGTATGAAAAACAACCAGTCTAAATCTGCC 1261

Qy     2542 ACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGTGATGGAGACTTGACAGAACAG 2601
Db     1262 AATTCAACCCTTCGGTTATTATCAGCGATGTTGGTTAGTGAGGGTGACCTGACAGAGCAA 1321

Qy     2602 GGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTTGCTGCTGGGAGTGCTATTGTG 2661
Db     1322 AAGAGGATCAGTAAATCTGATATGTCTCGCTTGCGATTAGCTGCTGGTAGTGCCATAATG 1381

Qy     2662 AAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACATTAGAACAATATCAGCTATGT 2721
Db     1382 AAGCTTGCTCAGGAACCTTGTTACCATGAAATTATTACCCCAGAACAGTTTCAGCTCTGT 1441

Qy     2722 GCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAAGTGTTTGCCCAGAAACTTCAC 2781
Db     1442 GCACTTGTTATTAATGATGAGTGTTACCAAGTAAGGCAGATATTTGCTCAGAAGCTGCAT 1501

Qy     2782 AAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATGGCAATCTGTGCCCTTTGTGCA 2841
Db     1502 AAGGCACTTGTGAAGTTACTGCTCCCATTGGAGTATATGGCGATCTTTGCCTTGTGTGCC 1561

Qy     2842 AAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAATGTTTGGTGAAAAATATAAAT 2901
Db     1562 AAAGATCCTGTGAAGGAGAGAAGAGCACACGCACGACAATGTTTACTGAAAAATATCAGT 1621

Qy     2902 GTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGTGAAAAATTATTGTCTCTTCTA 2961
Db     1622 ATACGCAGGGAATACATTAAGCAGAATCCTATGGCTACTGAGAAATTATTATCACTGTTG 1681

Qy     2962 CCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCACATGACCCAGATTATGTCAAA 3021
Db     1682 CCTGAATATGTAGTTCCATACATGATTCACCTGCTAGCCCATGATCCAGATTTTACAAGA 1741

Qy     3022 GTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGTCTTTGGTTTGTTCTGGAAATA 3081
Db     1742 TCACAAGATGTTGATCAGCTTCGTGATATCAAAGAGTGCCTATGGTTCATGCTTGAAGTT 1801

Qy     3082 TTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATCAGAAAGATGGTAGAAAATATT 3141
Db     1802 TTAATGACAAAGAATGAAAACAATAGCCATGCCTTTATGAAGAAGATGGCAGAGAACATC 1861

Qy     3142 AAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAAATGAATGAAAAACTGTACACT 3201
Db     1862 AAGTTAACCAGAGATGCCCAGTCTCCAGATGAATCCAAGACAAATGAAAAACTGTATACA 1921

Qy     3202 GTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGTACTACATACAGTTTGGAATCT 3261
Db     1922 GTATGTGATGTGGCTCTCTGTGTTATAAATAGTAAAAGTGCTTTGTGCAATGCAGATTCA 1981

Qy     3262 CCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAACCTGACAAGAATTTCAGTAAC 3321
Db     1982 CCAAAGGACCCAGTCCTCCCAATGAAATTTTTACACAACCTGAAAAGGACTTCTGTAAC 2041

Qy     3322 ACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTCACTCCTGGAAAACCTAAAACA 3381
Db     2042 GATAAGAGTTATATTTCAGAAGAGACAAGAGTACTTCTGTTAACAGGAAAGCCAAAGCCT 2101

Qy     3382 ACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCAGCAGGCAAGCAATCTCAGACC 3441
Db     2102 GCTGGAGTACTAGGTGCAGTAAATAAGCCTTTATCAGCAACGGGAAGGAAACCCTATGTT 2161

Qy     3442 AAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGCAGCTCAAATCCAAGCTCTCCT 3501
Db     2162 AGAAGCACTGGCACTGAGACTGGAAGCAATATTAATGTAAATTCAGAGCTGAACCCTTCA 2221

Qy     3502 GGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGATCACAGTGAAAATGAAGA 3557
Db     2222 ACCGGAAATCGATCAAGGGAACAGAGTTCAGAGGCAGCAGAAACTGGAGTTAGTGA 2277
```

RESULT 9
AAS44727
ID AAS44727 standard; DNA; 714 BP.
XX
AC AAS44727;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human full-length polynucleotide sequence #152.
XX
KW Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW nervous system disorder; inflammatory disorder; cell differentiation; ds;
KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;
KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;
KW immunostimulant; analgesic; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200164834-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04926.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
PR 17-JUN-2000; 2000US-0597707.
PR 14-JUL-2000; 2000US-0616807.
PR 19-SEP-2000; 2000US-0664641.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
PI Drmanac R;
XX
DR WPI; 2001-589862/66.
DR P-PSDB; AAU27827.
XX
PT Novel polypeptides and nucleic acids obtained from cDNA libraries
PT prepared from various human tissues, for diagnosis, treatment of
PT cancer, neurological, inflammatory disorders and for use in arrays for
PT detection -
XX
PS Claim 1; SEQ ID No 152; 153pp; English.
XX
CC Sequences AAS44576-AAS44919 represent full-length polynucleotides and
CC contig polynucleotides encoding polypeptides of the invention. The DNA
CC and protein sequences are useful for the treatment, diagnosis and
CC prevention of various types of disorder in a mammalian subject such as a

```
CC   human, dog, monkey, mouse, hamster or rat. The disorders include cancers
CC   such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such
CC   as multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC   diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC   disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC   chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC   Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC   disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC   bowel disease. The sequences exhibit activity relating to angiogenesis,
CC   cell proliferation, cell differentiation, stem cell growth factor,
CC   activin or inhibin. Therefore, they can be used to manipulate stem cells
CC   in culture to give rise to neuroepithelial cells that can be used to
CC   augment or replace cells damaged by illness, accidental damage or genetic
CC   disorders. The sequences may also be used for regeneration of bone,
CC   cartilage, tendons and ligaments and in tissue repair and burn healing.
CC   Note: Some sequences for this patent did not form part of the printed
CC   specification, but were obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 714 BP; 293 A; 131 C; 178 G; 112 T; 0 other;

  Query Match             12.0%;  Score 499;  DB 22;  Length 714;
  Best Local Similarity   100.0%;  Pred. No. 6.1e-113;
  Matches  499;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3674 AGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAACAGAAACCTAAAG 3733
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAACAGAAACCTAAAG 60

Qy       3734 GCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGTGGC 3793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGTGGC 120

Qy       3794 CTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGC 3853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGC 180

Qy       3854 CAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAG 3913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAG 240

Qy       3914 AGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAG 3973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAG 300

Qy       3974 AGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAAC 4033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAAC 360

Qy       4034 AGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAATCTAGTGCAATTG 4093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAATCTAGTGCAATTG 420

Qy       4094 AATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAAACGCCATCACCAT 4153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAAACGCCATCACCAT 480

Qy       4154 CACAACCAAAAAAAAATGT 4172
              |||||||||||||||||||
Db        481 CACAACCAAAAAAAAATGT 499


RESULT 10
ABK35358
ID   ABK35358 standard; cDNA; 2496 BP.
XX
AC   ABK35358;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Human cDNA encoding secreted protein #496.
XX
KW   Human; secreted protein; gene; ss; nutritional supplement; haemophilia;
KW   viral infection; bacterial infection; fungal infection; diabetes; asthma;
KW   autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;
KW   autoimmune thyroiditis; allergic reaction; neurodegenerative disease;
KW   Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;
KW   coagulation disorder; inflammatory disorder; Crohn's disease; incision;
KW   tissue regeneration; wound healing; burn; haematopoiesis;
KW   myeloid cell deficiency; lymphoid cell deficiency.
XX
OS   Homo sapiens.
XX
PN   WO200177288-A2.
XX
PD   18-OCT-2001.
XX
PF   29-MAR-2001; 2001WO-US10224.
XX
PR   06-APR-2000; 2000US-195582P.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Wong GG,  Clark HF,  Fechtel K,  Agostino MJ,  Howes SH,  Resnick RJ;
PI   Gulukota K,  Graham JR;
XX
DR   WPI; 2002-179321/23.
XX
PT   Five hundred and ninety two polynucleotides derived from a variety of
PT   human tissue sources which encode secreted proteins, useful for
PT   treating immune deficiencies and disorders such as autoimmune disorders
PT   -
XX
PS   Claim 1; Page 328-329; 372pp; English.
XX
CC   The invention relates to 592 polynucleotides which have been derived from
CC   a variety of human tissue sources and which encode novel secreted
CC   proteins. The polynucleotides can be used as probes for the
CC   identification and isolation of full length cDNA and genomic DNA. The
CC   polynucleotides and proteins can also be used as nutritional supplements.
CC   The proteins are useful in the treatment of various immune deficiencies
CC   and disorders such as viral infections, bacterial infections, fungal
CC   infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple
CC   sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions
CC   and conditions (e.g. asthma). They are also useful for treating
CC   neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
CC   disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
CC   inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
CC   useful for tissue regeneration, for wound healing and in the treatment of
CC   burns, incisions and ulcers. The proteins are also useful for regulating
CC   haematopoiesis and for treating myeloid or lymphoid cell deficiencies.
CC   Sequences ABK34863-ABK35454 represent polynucleotides of the invention.
XX
SQ   Sequence 2496 BP; 822 A; 430 C; 491 G; 753 T; 0 other;

  Query Match             11.8%;  Score 493;  DB 24;  Length 2496;
  Best Local Similarity   65.6%;  Pred. No. 3.3e-111;
  Matches  752;  Conservative    0;  Mismatches  390;  Indels    5;  Gaps    2;

Qy       2387 ATCTTCTCATGAATGATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATG 2446
              | |||     || |  |  |    ||||  ||||||     ||||| |||  ||||||||
Db         83 AGCTTGCACAGATTTTTGAGTCAACAGGTGAAAAGAATGGAAAACTGTGGTCTCCAGATG 142

Qy       2447 AAGAAGTATCTCCTGAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGC 2506
              |||| || || |||||   | | |  ||  | ||||| |||||| |  ||||  | ||||
Db        143 AAGAGGTTTCCCCTGAAGTACTAGCAAAGGTACAGGCAATTAAACTTCTGGTAAGGTGGC 202

Qy       2507 TACTTGGAATGAAAAATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAA 2566
              |  | || |||||||| || ||   ||||||| |  | ||| ||| |  | ||  || ||
Db        203 TGTTGGGTATGAAAAACAACCAGTCTAAATCTGCCAATTCAACCCTTCGGTTATTATCAG 262

Qy       2567 CAATATTGCATAGTGATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGT 2626
              | || |||  |||||| || ||| ||||||| ||   ||  || ||||||| | |||||||
Db        263 CGATGTTGGTTAGTGAGGGTGACCTGACAGAGCAAAAGAGGATCAGTAAATCTGATATGT 322

Qy       2627 CACGTCTGAGACTTGCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACC 2686
```

```
Db        323 CTCGCTTGCGATTAGCTGCTGGTAGTGCCATAATGAAGCTTGCTCAGGAACCTTGTTACC 382

Qy       2687 ATGAAATCATCACATTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCT 2746

Db        383 ATGAAATTATTACCCCAGAACAGTTTCAGCTCTGTGCACTTGTTATTAATGATGAGTGTT 442

Qy       2747 ATCAAGTAAGACAAGTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTC 2806

Db        443 ACCAAGTAAGGCAGATATTTGCTCAGAAGCTGCATAAGGCACTTGTGAAGTTACTGCTCC 502

Qy       2807 CACTTGAGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAG 2866

Db        503 CATTGGAGTATATGGCGATCTTTGCCTTGTGTGCCAAAGATCCTGTGAAGGAGAGAAGAG 562

Qy       2867 CTCATGCTAGGCAATGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGC 2926

Db        563 CACACGCACGACAATGTTTACTGAAAAATATCAGTATACGCAGGGAATACATTAAGCAGA 622

Qy       2927 ATGCAGCTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAA 2986

Db        623 ATCCTATGGCTACTGAGAAATTATTATCACTGTTGCCTGAATATGTAGTTCCATACATGA 682

Qy       2987 TTCACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAG 3046

Db        683 TTCACCTGCTAGCCCATGATCCAGATTTTACAAGATCACAAGATGTTGATCAGCTTCGTG 742

Qy       3047 ATGTTAAAGAATGTCTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACA 3106

Db        743 ATATCAAAGAGTGCCTATGGTTCATGCTTGAAGTTTTAATGACAAAGAATGAAAACAATA 802

Qy       3107 GTCACGCTTTTATCAGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGAC 3166

Db        803 GCCATGCCTTTATGAAGAAGATGGCAGAGAACATCAAGTTAACCAGAGATGCCCAGTCTC 862

Qy       3167 CAGATGATGCAAAAATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCA 3226

Db        863 CAGATGAATCCAAGACAAATGAAAAACTGTATACAGTATGTGATGTGGCTCTCTGTGTTA 922

Qy       3227 TCATGTCAAAGAGTACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTC 3286

Db        923 TAAATAGTAAAAGTGCTTTGTGCA--ATGCAGATTCCCAAAGGACCCAGCCTCCAATGAA 980

Qy       3287 GTTTCTTCACTCAACCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAA 3346

Db        981 ATTTTTACAC---ACCTGAAAAGGACTTCTGTAACGATAAGAGTTATATTTCAGAAGAGA 1037

Qy       3347 TGAAATCATTTTTCACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACA 3406

Db       1038 CAAGAGTACTTCTGTTAACAGGAAAGCCAAAGCCTGCTGGAGTACTAGGTGCAGTAAATA 1097

Qy       3407 AGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAA 3466

Db       1098 AGCCTTTATCAGCAACGGGAAGGAAACCCTATGTTAGAAGCACTGGCACTGAGACTGGAA 1157

Qy       3467 GCAATGCAAGCAGCAGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATA 3526

Db       1158 GCAATATTAATGTAAATTCAGAGCTGAACCCTTCAACCGGAAATCGATCAAGGGAACAGA 1217

Qy       3527 GTTCTGA 3533

Db       1218 GTTCAGA 1224
```

RESULT 11
AAK85948
ID   AAK85948 standard; DNA; 7943 BP.
XX
AC   AAK85948;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40760.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.

```
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 40760; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 7943 BP; 2567 A; 1175 C; 1468 G; 2733 T; 0 other;

  Query Match             10.5%;  Score 438;  DB 22;  Length 7943;
  Best Local Similarity   85.3%;  Pred. No. 2.1e-97;
  Matches  541;  Conservative    0;  Mismatches    0;  Indels   93;  Gaps    1;

Qy       3633 GGAGAAGCCTAGAGGCAGGAAAAAAACGCCCGTCACAGAACAGGAGGAGAAATTAGGTAT 3692
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGAGAAGCCTAGAGGCAGGAAAAAAACGCCCGTCACAGAACAGGAGGAGAAATTAGGTAT 60

Qy       3693 GGATGACTTGACTAAGTTGGTACAGGAACAGAAACCTAAAGGCAGTCAGCGAAGTCGGAA 3752
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGATGACTTGACTAAGTTGGTACAGGAACAGAAACCTAAAGGCAGTCAGCGAAGTCGGAA 120

Qy       3753 AAGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGTGGCCTGAGGAAAAGAGGCTCAA 3812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AAGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGTGGCCTGAGGAAAAGAGGCTCAA 180

Qy       3813 AGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCAAAAAAGGGTAAAAGAGG 3872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCAAAAAAGGGTAAAAGAGG 240

Qy       3873 CCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAGCCAACAATGAAAACTTC 3932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAGCCAACAATGAAAACTTC 300

Qy       3933 TAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGGAGGAAGAAGAAGA 3992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGGAGGAAGAAGAAGA 360

Qy       3993 AAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAACAGCACCGAGTGTCAAGGAG 4052
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAACAGCACCGAGTGTCAAGGAG 420

Qy       4053 AGCACAGCAG-------------------------------------------------- 4062
              ||||||||||
Db        421 AGCACAGCAGAGGTAAGCATGTGTAACTCTAAACTGCATCTGTTTCGTTACTATATTATA 480

Qy       4063 -------------------------------------------AGAGCAGAATCTCCTGA 4079
                                                         |||||||||||||||||
```

```
Db        481 AATCATAATTTGATGCTATCCACATTTGGGTCTTCCCCAAAGCAGAGCAGAATCTCCTGA 540

Qy       4080 ATCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAA 4139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAA 600

Qy       4140 AACGCCATCACCATCACAACCAAAAAAAAAATGTG 4173
              ||||||||||||||||||||||||||||||||||
Db        601 AACGCCATCACCATCACAACCAAAAAAAAAATGTG 634


RESULT 12
AAV87629
ID   AAV87629 standard; cDNA; 439 BP.
XX
AC   AAV87629;
XX
DT   12-FEB-1999  (first entry)
XX
DE   EST clone DY17.
XX
KW   Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW   tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW   chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW   receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS   Homo sapiens.
XX
PN   WO9845437-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06956.
XX
PR   10-APR-1997;   97US-0837312.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070078/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
XX
PS   Claim 1; Page 126; 641pp; English.
XX
CC   The present sequence represents an expressed sequence tag (EST), and is
CC   a polynucleotide of the invention. The polynucleotides of the invention
CC   are all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
XX
SQ   Sequence 439 BP; 150 A; 101 C; 85 G; 103 T; 0 other;

  Query Match             9.9%;  Score 411.4;  DB 20;  Length 439;
  Best Local Similarity   98.6%;  Pred. No. 2.1e-91;
  Matches  415;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       3216 CATGAATATCATCATGTCAAAGAGTACTACATACAGTTTGGAATCTCCTAAAGACCCGGT 3275
              ||||     |||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 CATGGCCTACATCATGTCAAAGAGTACTACATACAGTTTGGAATCTCCTAAAGACCCGGT 72

Qy       3276 ACTACCAGCTCGTTTCTTCACTCAACCTGACAAGAATTTCAGTAACACCAAAAATTATCT 3335
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 ACTACCAGCTCGTTTCTTCACTCAACCTGACAAGAATTTCAGTAACACCAAAAATTATCT 132

Qy       3336 GCCTCCTGAAATGAAATCATTTTTCACTCCTGGAAAACCTAAAACAACCAATGTTCTAGG 3395
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 GCCTCCTGAAATGAAATCATTTTTCACTCCTGGAAAACCTAAAACAACCAATGTTCTAGG 192

Qy       3396 AGCTGTTAACAAGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAAATCATCACGAAT 3455
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 AGCTGTTAACAAGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAAATCATCACGAAT 252

Qy       3456 GGAAACTGTAAGCAATGCAAGCAGCAGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGG 3515
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        253 GGAAACTGTAAGCAATGCAAGCAGCAGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGG 312

Qy       3516 GAGGCTTGATAGTTCTGAAATGGATCACAGTGAAAATGAAGATTACACAATGTCTTCACC 3575
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        313 GAGGCTTGATAGTTCTGAAATGGATCACAGTGAAAATGAAGATTACACAATGTCTTCACC 372

Qy       3576 TTTGCCGGGGAAAAAAAGTGACAAGAGAGACGACTCTGATCTTGTAAGGTCTGAATTGGA 3635
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        373 TTTGCCGGGGAAAAAAAGTGACAAGAGAGACGACTCTGATCTTGTAAGGTCTGAATCGGA 432

Qy       3636 G 3636
              |
Db        433 G 433


RESULT 13
ABL89669
ID   ABL89669 standard; cDNA; 461 BP.
XX
AC   ABL89669;
XX
DT   24-MAY-2002  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 231.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200190304-A2.
XX
PD   29-NOV-2001.
XX
PF   18-MAY-2001; 2001WO-US16450.
XX
PR   19-MAY-2000; 2000US-205515P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE,  Rosen CA;
XX
DR   WPI; 2002-122018/16.
DR   P-PSDB; ABB89260.
XX
PT   Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT   prevention of neural, immune system, muscular, reproductive,
PT   gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT   disorders -
XX
PS   Claim 4; SEQ ID NO 231; 2081pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC   (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
```

```
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 461 BP; 135 A; 84 C; 116 G; 121 T; 5 other;

  Query Match             8.5%;  Score 353.4;  DB 24;  Length 461;
  Best Local Similarity   98.9%;  Pred. No. 4.6e-77;
  Matches  354;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        104 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 163

Qy         61 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        164 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 223

Qy        121 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        224 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 283

Qy        181 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGGTAAAGATGTTCGCTTACTGGTA 240
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        284 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGATAAAGATGTTCGCTTACTGGTA 343

Qy        241 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        344 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 403

Qy        301 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATA 358
              ||||||||||| |||||||||||||||||||||||||||||||| |||||||| ||||
Db        404 GATAAACTAAANGATATATTTATGTTTATAACAAGACAGTTGAANGGGCTAGANGATA 461


RESULT 14
ABS56684
ID   ABS56684 standard; cDNA; 1104 BP.
XX
AC   ABS56684;
XX
DT   23-JAN-2003  (first entry)
XX
DE   Euchromosome fragile intelligence delay syndrome protein 20.02 cDNA.
XX
KW   Euchromosome fragile intelligence delay syndrome protein 20.02; human;
KW   euchromosome fragile intelligence delay syndrome; dementia; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             184..732
FT                   /*tag= a
FT                   /product= "euchromosome fragile intelligence delay
FT                             syndrome-associated protein, 20.02"
XX
PN   CN1351041-A.
XX
PD   29-MAY-2002.
XX
PF   26-OCT-2000; 2000CN-0125797.
XX
PR   26-OCT-2000; 2000CN-0125797.
XX
PA   (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX
PI   Mao Y,  Xie Y;
XX
DR   WPI; 2002-637117/69.
DR   P-PSDB; ABB84560.
XX
PT   Human euchromosome fragile intelligence delay syndrome associated
PT   protein 20.02 polypeptide, used to treat e.g. dementia -
XX
PS   Claim 6; Page 29-30 (disclosure); 33pp; Chinese.
XX
CC   This invention describes the novel human euchromosome fragile
CC   intelligence delay syndrome associated protein, 20.02. The polypeptide is
CC   used in treating diseases such as euchromosome fragile intelligence delay
CC   syndrome, and dementia. This sequence encodes the human euchromosome
CC   fragile intelligence delay syndrome-associated protein, 20.02 described
CC   in the disclosure of the invention.
XX
SQ   Sequence 1104 BP; 362 A; 188 C; 228 G; 326 T; 0 other;

  Query Match             8.4%;  Score 350.4;  DB 24;  Length 1104;
  Best Local Similarity   69.7%;  Pred. No. 3.7e-76;
  Matches  488;  Conservative    0;  Mismatches  211;  Indels    1;  Gaps    1;

Qy       1041 AGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACATGACCCTGAGGAAGCTATTAG 1100
              ||| || ||  | ||||| ||| | || || || |||||||| || || ||||||||| |
Db          1 AGCGAAGGATCTCACAGAATATTTAAAGGTTAGATCACATGATCCAGAAGAAGCTATTCG 60

Qy       1101 ACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAGGATATTCTTCTGGTCAATGA 1160
               |||||||| |||||  | ||| | |||||||| || | |||  | | | | || |||||
Db         61 TCATGATGTCATTGTTACTATAATAACAGCTGCCAAGAGGGACCTGCCT-TAGTAAATGA 119

Qy       1161 TCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAACGATGGAGAGTACGCAAAGA 1220
              |||  | |||   ||||| || || |||||| | || ||||| ||| ||||| | |||||
Db        120 TCAGCTGCTTGGCTTTGTAAGGGAAAGAACACTGGATAAACGGTGGCGAGTAAGAAAAGA 179

Qy       1221 AGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCTTTACAGTCAGCAGCTGGAAA 1280
              ||| |||||||| || || ||  |||||||||||||   | | ||    | ||| |||||
Db        180 AGCTATGATGGGTCTGGCTCAGCTTTATAAGAAATACTGTCTTCATGGTGAAGCAGGAAA 239

Qy       1281 AGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTGCTACATATATATTATCAAAA 1340
               || |||||| |  |  |    ||||| || |||||| | || ||||| |||||||| ||
Db        240 GGAAGCTGCAGAGAAAGTCAGCTGGATAAAGGACAAACTTCTGCATATTTATTATCAGAA 299

Qy       1341 TAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCTCAATACATGGTTCCTCACAA 1400
               || ||||| ||   | |  | || ||    ||||||||||| ||  | || || |||||
Db        300 CAGCATTGACGACAAATTGTTGGTAGAGAAAATCTTTGCTCAGTATCTTGTCCCCCACAA 359

Qy       1401 TTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTGTATGCCACACTGGATTTAAA 1460
                | |||||   |||  | |||||||||||||||||||| ||||| |   |||||  |||
Db        360 CCTGGAAACAGAAGAGAGAATGAAATGCTTATATTACTTATATGCTAGTTTGGATCCAAA 419

Qy       1461 TGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAATCTGCTCCGACATCAAGTAAA 1520
              |||||| |||||  | || ||||||||||| ||||| ||  |||| ||    || |||
Db        420 TGCTGTAAAAGCTCTCAACGAAATGTGGAAGTGTCAGAACATGCTTCGGAGCCATGTACG 479

Qy       1521 GGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCCAGTGTCAAGGCCATATTTTC 1580
               ||  |  | || |||   ||||| || | | |||| || |        ||||| |||
Db        480 CGAACTATTGGATTTGCACAAGCAGCCTACATCAGAGGCTAACTGTTCTGCCATGTTTGG 539

Qy       1581 AAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGTAAGGCTCAGGATTTCATGAA 1640
              |||| |||||   ||  |||  ||||| ||||| || || || || || |||||  ||||
Db        540 AAAACTGATGACCATAGCAAAGAATTTGCCTGACCCCGGGAAAGCACAAGATTTTGTGAA 599

Qy       1641 GAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGAAAGCAGTTAGAAGTACTTGT 1700
              |||||| |  |||||  | |  |||||||||||| |  |    ||||| ||  || |  |
Db        600 GAAATTTAACCAGGTTCTCGGCGATGATGAGAAACTTCGGTCTCAGTTGGAGTTATTAAT 659

Qy       1701 TAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1740
```

```
               ||| ||||| || || ||||| || || ||   |||||||
Db         660 TAGCCCAACCTGTTCTTGCAAACAAGCAGATATTTGTGTG 699


RESULT 15
AAZ15259
ID   AAZ15259 standard; cDNA; 738 BP.
XX
AC   AAZ15259;
XX
DT   12-OCT-1999  (first entry)
XX
DE   Human gene expression product cDNA sequence SEQ ID NO:2728.
XX
KW   Human; gene; gene expression product; diagnosis; therapy; probe;
KW   detection; mapping; tissue typing; profiling; forensic; cancer;
KW   genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9938972-A2.
XX
PD   05-AUG-1999.
XX
PF   28-JAN-1999;   99WO-US01619.
XX
PR   03-APR-1998;   98US-0080666.
PR   28-JAN-1998;   98US-0072910.
PR   24-FEB-1998;   98US-0075954.
PR   31-MAR-1998;   98US-0080114.
PR   03-APR-1998;   98US-0080515.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Crkvenjakov R,  Dickson M,  Drmanac R,  Drmanac S;
PI   Escobedo J,  Garcia PD,  Garcia V,  Giese K,  Innis MA;
PI   Jones WL,  Kassam A,  Kennedy GC,  Kita D,  Labat I;
PI   Lamson G,  Leshkowitz D,  Pot D,  Randazzo F,  Reinhard C;
PI   Stache-Crain B,  Sudduth-Klinger J,  Williams LT;
XX
DR   WPI; 1999-494092/41.
XX
PT   Novel human genes and their expression products which are
PT   differentially expressed in different cell types
XX
PS   Claim 1; Page 1324; 2479pp; English.
XX
CC   The present invention describes a library of human polynucleotides
CC   comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC   a method of detecting differentially expressed genes correlated with the
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample from a cell
CC   suspected of being cancerous, where the gene product is encoded by one
CC   of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC   polynucleotides can be used as a source of primers and probes, which can
CC   be used for a variety of purpose, e.g. detection of expression levels,
CC   mapping, tissue typing or profiling, forensics, genetic analysis and
CC   detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC   can be used for raising antibodies for experimental, diagnostic and
CC   therapeutic purposes. The polynucleotides may also be used to construct
CC   arrays for diagnostics (which may be used to determine function of an
CC   encoded protein); and to detect differences in expression levels between
CC   two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC   identify a genetic predisposition or susceptibility to a disease such as
CC   cancer). The polynucleotides of the invention are especially used in the
CC   diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC   and lung cancer. The polynucleotides can also be used to screen for
CC   peptide analogues and antagonists.
XX
SQ   Sequence 738 BP; 268 A; 145 C; 168 G; 150 T; 7 other;

  Query Match             8.3%;  Score 348;  DB 20;  Length 738;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-75;
  Matches  348;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3825 AGAAAATGAAGATGAACAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAA 3884
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          51 AGAAAATGAAGATGAACAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAA 110

Qy        3885 ACCTCTTGGTGGAGGTACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAG 3944
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         111 ACCTCTTGGTGGAGGTACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAG 170

Qy        3945 CAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGG 4004
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         171 CAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGG 230

Qy        4005 AAATACGGAACAGAAGTCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAG 4064
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         231 AAATACGGAACAGAAGTCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAG 290

Qy        4065 AGCAGAATCTCCTGAATCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACG 4124
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         291 AGCAGAATCTCCTGAATCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACG 350

Qy        4125 AGGAAGACCATCAAAAACGCCATCACCATCACAACCAAAAAAAAATGT 4172
               ||||||||||||||||||||||||||||||||||||||||||||||||
Db         351 AGGAAGACCATCAAAAACGCCATCACCATCACAACCAAAAAAAAATGT 398


Search completed: September 24, 2003, 10:05:30
Job time : 1019 secs
```

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2003  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:          September 24, 2003, 08:52:32 ; Search time 14865 Seconds
                                              (without alignments)
                                              11484.418 Million cell updates/sec

Title:           US-09-512-581B-3
Perfect score:   4173
Sequence:        1 atggctcattcaaagactag..........cacaaccaaaaaaaaatgtg 4173

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters:      5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                 1:  gb_ba:*
                 2:  gb_htg:*
                 3:  gb_in:*
                 4:  gb_om:*
                 5:  gb_ov:*
                 6:  gb_pat:*
                 7:  gb_ph:*
                 8:  gb_pl:*
                 9:  gb_pr:*
                 10:  gb_ro:*
                 11:  gb_sts:*
                 12:  gb_sy:*
                 13:  gb_un:*
                 14:  gb_vi:*
                 15:  em_ba:*
                 16:  em_fun:*
                 17:  em_hum:*
                 18:  em_in:*
                 19:  em_mu:*
                 20:  em_om:*
                 21:  em_or:*
                 22:  em_ov:*
                 23:  em_pat:*
                 24:  em_ph:*
                 25:  em_pl:*
                 26:  em_ro:*
                 27:  em_sts:*
                 28:  em_un:*
                 29:  em_vi:*
                 30:  em_htg_hum:*
                 31:  em_htg_inv:*
                 32:  em_htg_other:*
                 33:  em_htg_mus:*
                 34:  em_htg_pln:*
                 35:  em_htg_rod:*
                 36:  em_htg_mam:*
                 37:  em_htg_vrt:*
                 38:  em_sy:*
                 39:  em_htgo_hum:*
                 40:  em_htgo_mus:*
                 41:  em_htgo_other:*

    Pred. No. is the number of results predicted by chance to have a
    score greater than or equal to the score of the result being printed,
    and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4165.6	99.8	5309	9	AB023196	AB023196 Homo sapi
	2	4165.6	99.8	7444	9	HS26H231	AL137201 Novel hum
	3	4141	99.2	5253	9	HSU95825	U95825 Human andro
	4	3471.8	83.2	5317	10	AY102267	AY102267 Mus muscu
	5	2816.6	67.5	5944	10	AK122414	AK122414 Mus muscu
	6	1815.6	43.5	6744	9	AF294791	AF294791 Homo sapi
	7	1754.4	42.0	1852	9	HSU50533	U50533 Human BRCA2
	8	1464.6	35.1	4252	9	BC039256	BC039256 Homo sapi
	9	1041.4	25.0	5177	9	AB014548	AB014548 Homo sapi
	10	937.6	22.5	1976	10	BC021408	BC021408 Mus muscu
	11	932.2	22.3	2265	9	AK098331	AK098331 Homo sapi
	12	932.2	22.3	2705	9	BC041361	BC041361 Homo sapi
	13	447	10.7	2079	9	AK026889	AK026889 Homo sapi
	14	447	10.7	137246	9	HS49J10	Z84572 Human DNA s
	15	447	10.7	168487	9	AL138820	AL138820 Human DNA
c	16	443	10.6	72157	2	AC016449	AC016449 Homo sapi
	17	411.4	9.9	439	6	BD059747	BD059747 Secreted
c	18	389.8	9.3	54398	2	AC068224	AC068224 Homo sapi
	19	332.4	8.0	236508	2	AC111126	AC111126 Mus muscu
	20	292	7.0	295	6	BD026562	BD026562 Sequence
c	21	288.6	6.9	263479	2	AC125986	AC125986 Rattus no
	22	288	6.9	3779	9	BC009650	BC009650 Homo sapi
	23	232.2	5.6	772	6	BD145886	BD145886 Primer fo
	24	232.2	5.6	2212	6	BD159124	BD159124 Primer fo
	25	232.2	5.6	2212	9	AK021757	AK021757 Homo sapi
	26	208.8	5.0	227	6	AX494307	AX494307 Sequence
	27	207.4	5.0	54398	2	AC068224	AC068224 Homo sapi
	28	206.8	5.0	113704	9	HS267P19	Z75889 Human DNA s
	29	206.8	5.0	120652	9	AL353724	AL353724 Human DNA
	30	179.6	4.3	85592	2	AL512630	AL512630 Mus muscu
	31	179.6	4.3	179132	10	AL358892	AL358892 Mouse DNA
c	32	179.6	4.3	236508	2	AC111126	AC111126 Mus muscu
c	33	173.2	4.2	110000	2	AC094469_0	AC094469 Rattus no
	34	142.8	3.4	184321	2	AC110493	AC110493 Mus muscu
c	35	142.8	3.4	190058	2	AC112263	AC112263 Mus muscu
	36	142.8	3.4	298053	2	AC116574	AC116574 Mus muscu
c	37	142.6	3.4	233589	2	AC095342	AC095342 Rattus no
c	38	131	3.1	98856	9	AC113150	AC113150 Homo sapi
	39	131	3.1	145450	2	AC068352	AC068352 Homo sapi
	40	131	3.1	160962	2	AC022862	AC022862 Homo sapi
	41	131	3.1	167671	2	AC090081	AC090081 Homo sapi
c	42	124	3.0	69740	2	AC087748	AC087748 Homo sapi
c	43	121.4	2.9	145450	2	AC068352	AC068352 Homo sapi
	44	121.4	2.9	162775	2	AC023006	AC023006 Homo sapi
	45	119.2	2.9	173613	3	AC007475	AC007475 Drosophil

ALIGNMENTS

```
RESULT 1
AB023196
LOCUS       AB023196                5309 bp    mRNA    linear   PRI 10-MAY-2002
DEFINITION  Homo sapiens mRNA for KIAA0979 protein, partial cds.
ACCESSION   AB023196
VERSION     AB023196.2  GI:20521717
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirosawa,M.,
            Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
```

```
            XIII. The complete sequences of 100 new cDNA clones from brain
            which code for large proteins in vitro
  JOURNAL   DNA Res. 6 (1), 63-70 (1999)
  MEDLINE   99246063
   PUBMED   10231032
REFERENCE   2  (bases 1 to 5309)
  AUTHORS   Ohara,O., Nagase,T. and Kikuno,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-1999) Osamu Ohara, Kazusa DNA Research Institute,
            Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
            292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913,
            Fax:+81-438-52-3914)
COMMENT     On May 9, 2002 this sequence version replaced gi:4589601.
FEATURES             Location/Qualifiers
     source          1. .5309
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="hj07056s1"
                     /sex="male"
                     /tissue_type="brain"
                     /clone_lib="pBluescriptII SK plus"
                     /dev_stage="adult"
                     /note="This sequence was obtained bysubcloning of the DNA
                     fragments derived from two cDNA clones (1 - 722 was
                     derived from fj09522 and 723 - 5309 was derived from
                     hj07056)."
     gene            1. .5309
                     /gene="KIAA0979"
     CDS             <1. .4452
                     /gene="KIAA0979"
                     /codon_start=1
                     /product="KIAA0979 protein"
                     /protein_id="BAA76823.2"
                     /db_xref="GI:20521718"
                     /translation="GGGRRRLDAAAAADPGEPRSERSSESATRRGRNISVMAHSKTRT
                     NDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLALHLASD
                     FFLKHPDKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQ
                     FNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMS
                     SIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITNFFNQV
                     LMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAK
                     MFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLK
                     VRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQ
                     IYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTE
                     RMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVM
                     VITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLG
                     NPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAI
                     RAGLELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFP
                     HIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEH
                     LITPLVTIGHIALLAPDQFAAPLKSLVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSP
                     ETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRL
                     RLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLP
                     LEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPY
                     TIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKD
                     AQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKN
                     YLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPG
                     RIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQE
                     EKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESDEQQWPEEKRLKEDILENEDEQNSP
                     PKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEERQSGNTEQKSK
                     SKQHRVSRRAQQRAESPESSAIESTQSTPQKGRGRPSKTPSPSQPKKNVRVGRSKQAA
                     TKENDSSEEVDVFQGSSPVDDIPQEETEEEEVSTVNVRRRSAKRERR"
BASE COUNT     1774 a    955 c   1115 g   1465 t
ORIGIN

  Query Match             99.8%;  Score 4165.6;  DB 9;  Length 5309;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 4168;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        109 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 168

Qy         61 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        169 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 228

Qy        121 ACTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        229 ACTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 288

Qy        181 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGGTAAAGATGTTCGCTTACTGGTA 240
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        289 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGATAAAGATGTTCGCTTACTGGTA 348

Qy        241 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        349 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 408

Qy        301 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        409 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 468

Qy        361 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        469 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 528

Qy        421 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        529 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 588

Qy        481 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        589 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 648

Qy        541 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        649 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 708

Qy        601 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        709 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 768

Qy        661 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCACTTTTTTTAATCAGGTT 720
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        769 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTTTTAATCAGGTT 828

Qy        721 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        829 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 888

Qy        781 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        889 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 948

Qy        841 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        949 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 1008

Qy        901 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1009 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 1068

Qy        961 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1069 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1128

Qy       1021 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1129 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1188

Qy       1081 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1189 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1248

Qy       1141 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     1249 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1308

Qy     1201 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1309 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1368

Qy     1261 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1369 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1428

Qy     1321 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1429 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1488

Qy     1381 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1489 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1548

Qy     1441 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1549 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1608

Qy     1501 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1609 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1668

Qy     1561 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1669 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1728

Qy     1621 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1729 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1788

Qy     1681 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1789 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1848

Qy     1741 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1849 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1908

Qy     1801 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1909 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 1968

Qy     1861 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1969 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 2028

Qy     1921 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2029 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 2088

Qy     1981 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2089 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2148

Qy     2041 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2149 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2208

Qy     2101 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2209 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2268

Qy     2161 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2269 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2328

Qy     2221 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2329 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2388

Qy     2281 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2389 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2448

Qy     2341 TTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2400
            ||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||
Db     2449 TTTGCTGCTCCTTTGAAATCTTTGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2508

Qy     2401 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2509 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2568

Qy     2461 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2569 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2628

Qy     2521 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2629 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2688

Qy     2581 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2689 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2748

Qy     2641 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2749 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2808

Qy     2701 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2809 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2868

Qy     2761 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2820
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2869 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2928

Qy     2821 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 2880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2929 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 2988

Qy     2881 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 2940
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2989 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 3048

Qy     2941 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3049 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3108

Qy     3001 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3060
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3109 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3168

Qy     3061 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3169 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3228

Qy     3121 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3229 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3288

Qy     3181 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3289 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3348

Qy     3241 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3349 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3408

Qy     3301 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3409 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3468
```

```
Qy        3361 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3469 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3528

Qy        3421 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3529 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3588

Qy        3481 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3589 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3648

Qy        3541 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3649 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3708

Qy        3601 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3709 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3768

Qy        3661 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3769 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3828

Qy        3721 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3829 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3888

Qy        3781 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3889 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3948

Qy        3841 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 3900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3949 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 4008

Qy        3901 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 3960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4009 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 4068

Qy        3961 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4069 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4128

Qy        4021 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4129 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4188

Qy        4081 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        4189 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4248

Qy        4141 ACGCCATCACCATCACAACCAAAAAAAAATGT 4172
               ||||||||||||||||||||||||||||||||
Db        4249 ACGCCATCACCATCACAACCAAAAAAAAATGT 4280


RESULT 2
HS26H231
LOCUS       HS26H231                7444 bp    mRNA    linear   PRI 25-JAN-2000
DEFINITION  Novel human gene mapping to chomosome 13.
ACCESSION   AL137201
VERSION     AL137201.1  GI:6759511
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 7444)
  AUTHORS   Rhodes,S. and Huckle,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-2000) E-mail contact: humquery@sanger.ac.uk
COMMENT     This cDNA sequence was assembled from public domain ESTs and single
            pass sequencing reads from expressed DNA templates, aligned to the
            genomic DNA sequence from the bacterial clones 26H23 (Z84467),
            267P19 (Z75889) and 49J10( Z84572).
            The EST sequences listed match this sequence with an identity of at
            least 95% between the coordinates shown.
            Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13/ Experimentally determined gene
            Sanger Centre name : 26H23.C13.1.
FEATURES             Location/Qualifiers
     source          1..7444
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /map="13q12.3"
     exon            1..107
                     /number=1
     exon            108..234
                     /number=2
     CDS             127..4470
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB69911.1"
                     /db_xref="GI:6759512"
                     /translation="MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMD
                     MDQDSEEEKELYLNLALHLASDFFLKHPDKDVRLLVACCLADIFRIYAPEAPYTSPDK
                     LKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRT
                     LFSVINNGHNQKVHMHMVDLMSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLA
                     KALLKRTAQAIEPYITNFFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQ
                     LEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECV
                     KFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNF
                     VRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSID
                     DRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKD
                     LLDLIKQPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVL
                     VSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQ
                     VNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDE
                     KVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSK
                     ETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPLKSLVATFIVKDLL
                     MNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTT
                     ILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDE
                     CYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREY
                     LKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMA
                     KNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESP
                     KDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQ
                     TKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKRDDS
                     DLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESDEQ
                     QWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSG
                     PPAPEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESPESSAIESTQSTPQKGRGRP
                     SKTPSPSQPKKNVRVGRSKQAATKENDSSEEVDVFQGSSPVDDIPQEETEEEEVSTVN
                     VRRRSAKRERR"
     exon            235..438
                     /number=3
     exon            439..525
                     /number=4
     exon            526..623
                     /number=5
     exon            624..750
                     /number=6
     exon            751..831
                     /number=7
     exon            832..972
                     /number=8
     exon            973..1088
                     /number=9
     exon            1089..1183
                     /number=10
     exon            1184..1329
                     /number=11
     exon            1330..1481
                     /number=12
     exon            1482..1595
                     /number=13
     exon            1596..1677
```

```
                     /number=14
     exon            1678..1726
                     /number=15
     exon            1727..1866
                     /number=16
     misc_feature    1866..2046
                     /note="matches EST AA687846 from clone IMAGE:1219676"
     misc_feature    1866..1941
                     /note="matches EST AA687776 from clone IMAGE:1219628"
     exon            1867..1982
                     /number=17
     exon            1983..2088
                     /number=18
     exon            2089..2249
                     /number=19
     exon            2250..2373
                     /number=20
     exon            2374..2532
                     /number=21
     exon            2533..2601
                     /number=22
     exon            2602..2738
                     /number=23
     exon            2739..2862
                     /number=24
     exon            2863..3067
                     /number=25
     misc_feature    complement(2912..3302)
                     /note="matches EST AA630616 from clone 856164"
     exon            3068..3182
                     /number=26
     exon            3183..3315
                     /number=27
     misc_feature    3302..3572
                     /note="matches EST AA333118"
     exon            3316..3435
                     /number=28
     misc_feature    3349..3818
                     /note="matches EST AA282190 from clone IMAGE:704631"
     exon            3436..3498
                     /number=29
     exon            3499..3644
                     /number=30
     misc_feature    complement(3586..4065)
                     /note="matches EST AI655429 from clone IMAGE:2239331"
     exon            3645..3750
                     /number=31
     exon            3751..4190
                     /number=32
     misc_feature    complement(3864..4286)
                     /note="matches EST AI680124 from clone IMAGE:2264521"
     misc_feature    4001..4119
                     /note="matches EST N83765 from clone KK2745"
     misc_feature    4135..4450
                     /note="matches EST F12354 from clone c-38h11."
     exon            4191..4298
                     /number=33
     misc_feature    4255..4307
                     /note="matches EST T64867 from clone 22264"
     exon            4299..4434
                     /number=34
     misc_feature    join(4398..4533,4518..4660,4659..4675,4674..4693)
                     /note="matches EST H44781 from clone 188032"
     exon            4435..7444
                     /number=35
     misc_feature    4537..5022
                     /note="matches EST AI681422 from clone IMAGE:2272632"
     misc_feature    4537..4892
                     /note="matches EST AA911125 from clone IMAGE:1416862"
     misc_feature    4538..4969
                     /note="matches EST AI275866 from clone IMAGE:1877482"
     misc_feature    4538..4947
                     /note="matches EST AA987361 from clone IMAGE:1603350"
```

```
     misc_feature    4538..4943
                     /note="matches EST AI299646 from clone IMAGE:1898010"
     misc_feature    4562..5003
                     /note="matches EST AA447433 from clone 784555"
     misc_feature    4571..4854
                     /note="matches EST H93424 from clone 220700"
     misc_feature    join(4573..4649,4650..4815)
                     /note="matches EST W20070 from clone 306049"
     misc_feature    complement(4629..4937)
                     /note="matches EST AW089775 from clone IMAGE:2593146"
     misc_feature    4746..4994
                     /note="matches EST AA747568 from clone IMAGE:1269696"
     misc_feature    complement(4803..5332)
                     /note="matches EST AI911784 from clone IMAGE:2329873"
     misc_feature    complement(4828..5296)
                     /note="matches EST AA563884 from clone IMAGE:1013990"
     misc_feature    4861..5243
                     /note="matches EST D20082 from clone pm1283."
     misc_feature    join(4872..4921,5281..5371)
                     /note="matches EST AA177347 from clone 621810"
     misc_feature    complement(join(4872..4921,5281..5332))
                     /note="matches EST AW212954 from clone IMAGE:2647801"
     misc_feature    4872..4921
                     /note="matches EST AA154932 from clone 599194"
     misc_feature    complement(4872..4921)
                     /note="matches EST AW048929 from clone
                     UI-M-BH1-amo-c-09-0-UI"
     misc_feature    complement(4876..5333)
                     /note="matches EST AI016896 from clone IMAGE:1627886"
     misc_feature    complement(join(4922..4975,5281..5332))
                     /note="matches EST AI012649 from clone RPLBA02"
     misc_feature    complement(4922..4975)
                     /note="matches EST AA899586 from clone
                     UI-R-E0-cy-e-02-0-UI
                     matches EST AI172172 from clone RMUBV82"
     misc_feature    complement(4931..5332)
                     /note="matches EST AA625961 from clone 745478"
     misc_feature    complement(4951..5332)
                     /note="matches EST AI140790 from clone IMAGE:1566222"
     misc_feature    complement(5021..5309)

  Query Match             99.8%;  Score 4165.6;  DB 9;  Length 7444;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 4168;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 186

Qy         61 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 246

Qy        121 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 306

Qy        181 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGGTAAAGATGTTCGCTTACTGGTA 240
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        307 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGATAAAGATGTTCGCTTACTGGTA 366

Qy        241 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 426

Qy        301 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 486

Qy        361 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        487 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 546
```

```
Qy        421 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        547 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 606

Qy        481 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        607 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 666

Qy        541 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        667 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 726

Qy        601 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        727 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 786

Qy        661 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCACTTTTTTTAATCAGGTT 720
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        787 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTTTTAATCAGGTT 846

Qy        721 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        847 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 906

Qy        781 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        907 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 966

Qy        841 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        967 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 1026

Qy        901 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1027 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 1086

Qy        961 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1087 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1146

Qy       1021 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1147 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1206

Qy       1081 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1207 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1266

Qy       1141 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1267 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1326

Qy       1201 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1327 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1386

Qy       1261 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1387 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1446

Qy       1321 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1447 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1506

Qy       1381 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1507 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1566

Qy       1441 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1567 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1626

Qy       1501 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1627 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1686

Qy       1561 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1687 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1746

Qy       1621 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1747 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1806

Qy       1681 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1807 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1866

Qy       1741 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1867 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1926

Qy       1801 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1927 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 1986

Qy       1861 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1987 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 2046

Qy       1921 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2047 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 2106

Qy       1981 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2107 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2166

Qy       2041 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2167 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2226

Qy       2101 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2227 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2286

Qy       2161 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2287 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2346

Qy       2221 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2347 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2406

Qy       2281 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2407 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2466

Qy       2341 TTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2400
              ||||||||||||| ||||||||| ||||||||||||||||||||||||||||||||||||
Db       2467 TTTGCTGCTCCTTTGAAATCTTTGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2526

Qy       2401 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2527 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2586

Qy       2461 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2587 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2646

Qy       2521 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACACAACAATATTGCATAGT 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2647 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACACAACAATATTGCATAGT 2706

Qy       2581 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   2707 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2766

Qy   2641 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2767 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2826

Qy   2701 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2827 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2886

Qy   2761 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2887 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2946

Qy   2821 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2947 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 3006

Qy   2881 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3007 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 3066

Qy   2941 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3067 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3126

Qy   3001 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3127 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3186

Qy   3061 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3187 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3246

Qy   3121 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3247 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3306

Qy   3181 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3307 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3366

Qy   3241 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3367 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3426

Qy   3301 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3427 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3486

Qy   3361 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3487 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3546

Qy   3421 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3547 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3606

Qy   3481 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3607 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3666

Qy   3541 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3667 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3726

Qy   3601 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3727 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3786

Qy   3661 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3787 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3846

Qy   3721 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3847 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3906

Qy   3781 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3907 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3966

Qy   3841 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3967 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 4026

Qy   3901 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4027 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 4086

Qy   3961 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4087 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4146

Qy   4021 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4147 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4206

Qy   4081 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4207 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4266

Qy   4141 ACGCCATCACCATCACAACCAAAAAAAAATGT 4172
          ||||||||||||||||||||||||||||||||
Db   4267 ACGCCATCACCATCACAACCAAAAAAAAATGT 4298
```

```
RESULT 3
HSU95825
LOCUS       HSU95825                5253 bp    mRNA    linear   PRI 05-APR-1999
DEFINITION  Human androgen-induced prostate proliferative shutoff associated
            protein (AS3) mRNA, complete cds.
ACCESSION   U95825
VERSION     U95825.2  GI:4559409
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5253)
  AUTHORS   Geck,P., Szelei,J., Jimenez,J., Soto,A.M. and Sonnenschein,C.
  TITLE     Androgen-induced proliferative shutoff in prostate cancer cells
  JOURNAL   Proc. Annu. Meet. Am. Assoc. Cancer Res. 37, 223-223 (1996)
REFERENCE   2  (bases 1 to 5253)
  AUTHORS   Geck,P., Szelei,J., Jimenez,J., Sonnenschein,C. and Soto,A.M.
  TITLE     Early gene expression during androgen-induced inhibition of
            proliferation of prostate cancer cells: a new suppressor candidate
            on chromosome 13, in the BRCA2-Rb1 locus
  JOURNAL   J. Steroid Biochem. Mol. Biol. 68 (1-2), 41-50 (1999)
  MEDLINE   99229875
   PUBMED   10215036
REFERENCE   3  (bases 1 to 5253)
  AUTHORS   Geck,P., Szelei,J., Jimenez,J., Sonnenschein,C. and Soto,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-1997) Anatomy and Cell Biology, Tufts University
            Medical School, 136 Harrison Avenue, Boston, MA 02111, USA
REFERENCE   4  (bases 1 to 5253)
  AUTHORS   Geck,P., Szelei,J., Jimenez,J., Sonnenschein,C. and Soto,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-1999) Anatomy and Cell Biology, Tufts University
            Medical School, 136 Harrison Avenue, Boston, MA 02111, USA
  REMARK    Sequence update by submitter
COMMENT     On Apr 5, 1999 this sequence version replaced gi:4539617.
FEATURES             Location/Qualifiers
     source          1..5253
                     /organism="Homo sapiens"
```

```
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /map="13q12-q13"
                     /cell_line="LNCaP"
                     /cell_type="cancer cells"
                     /tissue_type="prostate"
                     /note="exon 1 maps to P1 Artificial Chromosome PAC26H23;
                     exons 2-24 map to cosmid 267p19; exons 25-34 map to P1
                     Artificial Chromosome PAC 49J10"
     gene            1. .5253
                     /gene="AS3"
     exon            1. .46
                     /gene="AS3"
                     /number=1
     exon            47. .173
                     /gene="AS3"
                     /number=2
     CDS             66. .4241
                     /gene="AS3"
                     /codon_start=1
                     /product="androgen-induced prostate proliferative shutoff
                     associated protein"
                     /protein_id="AAD22134.2"
                     /db_xref="GI:4559410"
                     /translation="MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMD
                     MDQDSEEEKELYLNLALHLASDFFLKHPDKDVRLLVACCLADIFRIYAPEAPYTSPDK
                     LKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRT
                     LFSVINNGHNQKVHMHMVDLMSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLA
                     KALLKRTAQAIEPYITNFFNQVLMLGKTSISDLSEHVSDLILELYNIDSHLLLSVLPQ
                     LEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECV
                     KFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNF
                     VGERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSID
                     DRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKD
                     LLDLIKQPKTDASVKAIFSKVMVITRSLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVL
                     VSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQ
                     VNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDE
                     KVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSK
                     EAQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPLKSLVATFIVKDLL
                     MNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTT
                     ILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDE
                     CYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREY
                     LKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMA
                     KNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESP
                     KDPVLPARFFTQPDKNFSNTKNYLPPGMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQ
                     TKSSRMETVGNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKRSDKRDDS
                     DLVRSELEKPRGRKKTPVTEREEKLGMDDLTKLVQEQRPKGSQRSRKRGHTASESDEQ
                     QWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSG
                     PPAPEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESSESSAIESTQSTPQKGRGRP
                     SKTPSPSQPKKNV"
     exon            174. .377
                     /gene="AS3"
                     /number=3
     exon            378. .464
                     /gene="AS3"
                     /number=4
     exon            465. .562
                     /gene="AS3"
                     /number=5
     exon            563. .689
                     /gene="AS3"
                     /number=6
     exon            690. .770
                     /gene="AS3"
                     /number=7
     exon            771. .911
                     /gene="AS3"
                     /number=8
     exon            912. .1027
                     /gene="AS3"
                     /number=9
     exon            1028. .1122
                     /gene="AS3"
                     /number=10
     exon            1123. .1268
                     /gene="AS3"
                     /number=11
     exon            1269. .1420
                     /gene="AS3"
                     /number=12
     exon            1421. .1534
                     /gene="AS3"
                     /number=13
     exon            1535. .1616
                     /gene="AS3"
                     /number=14
     exon            1617. .1665
                     /gene="AS3"
                     /number=15
     exon            1666. .1805
                     /gene="AS3"
                     /number=16
     exon            1806. .1921
                     /gene="AS3"
                     /number=17
     exon            1922. .2027
                     /gene="AS3"
                     /number=18
     exon            2028. .2188
                     /gene="AS3"
                     /number=19
     exon            2189. .2312
                     /gene="AS3"
                     /number=20
     exon            2313. .2471
                     /gene="AS3"
                     /number=21
     exon            2472. .2540
                     /gene="AS3"
                     /number=22
     exon            2541. .2677
                     /gene="AS3"
                     /number=23
     exon            2678. .2801
                     /gene="AS3"
                     /number=24
     exon            2802. .3006
                     /gene="AS3"
                     /number=25
     exon            3007. .3121
                     /gene="AS3"
                     /number=26
     exon            3122. .3254
                     /gene="AS3"
                     /number=27
     exon            3255. .3374
                     /gene="AS3"
                     /number=28
     exon            3375. .3437
                     /gene="AS3"
                     /number=29
     exon            3438. .3583
                     /gene="AS3"
                     /number=30
     exon            3584. .3689
                     /gene="AS3"
                     /number=31
     exon            3690. .4129
                     /gene="AS3"
                     /number=32
     exon            4130. .4354
                     /gene="AS3"
                     /number=33
     exon            4355. .4810
                     /gene="AS3"
                     /number=34
```

```
BASE COUNT      1755 a    944 c    1074 g    1480 t
ORIGIN

  Query Match             99.2%;  Score 4141;  DB 9;  Length 5253;
  Best Local Similarity   99.5%;  Pred. No. 0;
  Matches 4153;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       1 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 125

Qy      61 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     126 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 185

Qy     121 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     186 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 245

Qy     181 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGGTAAAGATGTTCGCTTACTGGTA 240
           ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     246 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGATAAAGATGTTCGCTTACTGGTA 305

Qy     241 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     306 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 365

Qy     301 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     366 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 425

Qy     361 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 420
           ||||||||||| |||||||||||||||||||| |||||||||||||||||||||||||||
Db     426 AAGAGCCCACAGTTCAATAGGTATTTTTATTTGCTTGAGAACATTGCTTGGGTCAAGTCA 485

Qy     421 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     486 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 545

Qy     481 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     546 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 605

Qy     541 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     606 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 665

Qy     601 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     666 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 725

Qy     661 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCACTTTTTTTAATCAGGTT 720
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     726 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTTTTAATCAGGTT 785

Qy     721 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 780
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db     786 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTCTGACTTAATTTTG 845

Qy     781 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     846 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 905

Qy     841 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 900
           |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     906 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCGAAAATGTTT 965

Qy     901 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     966 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 1025

Qy     961 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1026 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1085

Qy    1021 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1086 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1145

Qy    1081 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1146 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1205

Qy    1141 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1200
           ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    1206 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGGGAGAGAGAACATTAGACAAA 1265

Qy    1201 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1266 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1325

Qy    1261 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1326 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1385

Qy    1321 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1386 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1445

Qy    1381 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1446 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1505

Qy    1441 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1506 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1565

Qy    1501 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1566 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1625

Qy    1561 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1620
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    1626 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAGTTTACCTGATCCTGGT 1685

Qy    1621 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1686 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1745

Qy    1681 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1746 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1805

Qy    1741 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1806 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1865

Qy    1801 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1866 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 1925

Qy    1861 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1926 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 1985

Qy    1921 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1986 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 2045

Qy    1981 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2040
           |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db    2046 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCCTGTCTGAAAATGGAT 2105

Qy    2041 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2106 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2165
```

```
Qy       2101 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2166 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2225

Qy       2161 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2226 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2285

Qy       2221 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2280
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2286 GAGGCCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2345

Qy       2281 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2346 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2405

Qy       2341 TTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2400
              ||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||||
Db       2406 TTTGCTGCTCCTTTGAAATCTTTGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2465

Qy       2401 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2466 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2525

Qy       2461 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2526 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2585

Qy       2521 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2586 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2645

Qy       2581 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2646 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2705

Qy       2641 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2706 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2765

Qy       2701 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2766 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2825

Qy       2761 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2826 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2885

Qy       2821 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2886 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 2945

Qy       2881 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2946 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 3005

Qy       2941 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3006 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3065

Qy       3001 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3066 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3125

Qy       3061 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3126 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3185

Qy       3121 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3180
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db       3186 AGAAAGATGGTAGAAAATATTAAGCAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3245

Qy       3181 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3246 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3305

Qy       3241 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3306 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3365

Qy       3301 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3360
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       3366 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGGAATGAAATCATTTTTC 3425

Qy       3361 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3426 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3485

Qy       3421 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3480
              ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       3486 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAGGCAATGCAAGCAGC 3545

Qy       3481 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3546 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3605

Qy       3541 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       3606 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAGAAGTGACAAG 3665

Qy       3601 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3660
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       3666 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGAAAAAAAACG 3725

Qy       3661 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3720
              ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       3726 CCCGTCACAGAACGGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3785

Qy       3721 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3780
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3786 CAGAGACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3845

Qy       3781 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3846 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3905

Qy       3841 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3906 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 3965

Qy       3901 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 4025

Qy       3961 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4026 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4085

Qy       4021 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db       4086 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTTCTGAA 4145

Qy       4081 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4146 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4205

Qy       4141 ACGCCATCACCATCACAACCAAAAAAAAATGTG 4173
              |||||||||||||||||||||||||||||||||
Db       4206 ACGCCATCACCATCACAACCAAAAAAAAATGTG 4238
```

RESULT 4
AY102267
LOCUS   AY102267   5317 bp   mRNA   linear   ROD 01-JUN-2003
DEFINITION Mus musculus androgen-induced prostate proliferative shutoff

```
            associated protein AS3 mRNA, complete cds.
ACCESSION   AY102267
VERSION     AY102267.1  GI:31321922
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 5317)
  AUTHORS   Geck,P., Maffini,M., Sonnenschein,C. and Soto,A.M.
  TITLE     The AS3 proliferative arrest gene has an ancient eukaryotic
            heritage and shows highly conserved functional domains in mice
  JOURNAL   Proc. Annu. Meet. Am. Assoc. Cancer Res. 43, 987 (2002)
REFERENCE   2  (bases 1 to 5317)
  AUTHORS   Geck,P., Maffini,M., Liang,S.L., Sonnenschein,C. and Soto,A.M.
  TITLE     Extreme domain conservation and expression of the mouse AS3
            proliferative arrest protein
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5317)
  AUTHORS   Geck,P., Maffini,M., Liang,S.L., Sonnenschein,C. and Soto,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2002) Anatomy and Cell Biology, Tufts University
            School of Medicine, 136 Harrison Avenue, Boston, MA 02111, USA
FEATURES             Location/Qualifiers
     source          1..5317
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /map="84cM; G2; between BRCA2 and KL"
                     /sex="male"
                     /tissue_type="brain"
     CDS             147..4487
                     /codon_start=1
                     /product="androgen-induced prostate proliferative shutoff
                     associated protein AS3"
                     /protein_id="AAM52216.1"
                     /db_xref="GI:31321923"
                     /translation="MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMD
                     MDQDSEEEKELYLNLALHLASDFFLKHPDKDVRLLVACCLADIFRIYAPEAPYTSPDK
                     LKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRT
                     LFSVINNGHNQKVHMHMVDLMSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLA
                     KALLKRTAQAIEPYITNFFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQ
                     LEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECV
                     KFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNF
                     VRERTLDKRWRVRKEAMMGLAQIYKKYSLQSAAGKDAAKQISWVKDKLLHIYYQNSID
                     DRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMRKCQNLLRHQVKD
                     LLDLIKQPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEAL
                     VSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQ
                     VNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDE
                     KVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSK
                     ETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPLKSLVATFIVKDLL
                     MNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTT
                     ILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDE
                     CYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNTTVRREY
                     LKQHASVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMA
                     KNENNSHAFIRKMVENIKQTKDAQGPDDTKMNEKLYTVCDVAMNIIMSKSTTYSLESP
                     KDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTANVLGAVNKPLSSAGKQSQ
                     TKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKREDP
                     DLSELEKPRSRKKAPVTDPEEKLGMDDLTKLVQEQKPKGSQRGRKRGRTASDSDEQQW
                     PEEKRHKEELLENEDEQNSPPKKGKRGRPPKPLGGGTSKEEPTMKTSKKGNKKKLVPP
                     VVDDDEEEERQIGNTEHKSKSKQHRTSKRAQQRAESPETSAVESTQSTPQKGRGRPSK
                     APSPSQPPKKIRVGRSKQVATKENDSSEEMDVLQASSPVSDDTTQEGAEEEDISVGNV
                     RRRSSKRERR"
BASE COUNT     1676 a   1100 c   1199 g   1342 t
ORIGIN

  Query Match             83.2%;  Score 3471.8;  DB 10;  Length 5317;
  Best Local Similarity   89.7%;  Pred. No. 0;
  Matches 3741;  Conservative    0;  Mismatches  422;  Indels    6;  Gaps    1;

Qy          1 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 60
              |||||||||||||||| |||||||| ||||| |||||||| || || ||||| ||||||
Db        147 ATGGCTCATTCAAAGACAAGGACCAACGATGGGAAAATTACTTACCCTCCTGGAGTCAAG 206

Qy         61 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 120
              ||||| ||||||||||| |||||||||||||||||| || ||||||||||||| |||
Db        207 GAAATCTCAGATAAAATCTCTAAAGAGGAGATGGTGAGGCGGTTAAAGATGGTTGTAAAA 266

Qy        121 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 180
              ||||| ||||| ||||||||||||||||||| |||||||| |||||| ||||||||||
Db        267 ACTTTCATGGACATGGACCAGGACTCTGAAGAGGAAAAGGAACTTTATCTAAACCTAGCT 326

Qy        181 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGGTAAAGATGTTCGCTTACTGGTA 240
              ||||||||||||| || || || |||||||||||||| |||||||||||| ||||||||
Db        327 TTACATCTTGCTTCTGACTTCTTCCTCAAGCATCCTGATAAAGATGTTCGTTTACTGGTG 386

Qy        241 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 300
              || |||||||||||||||||||||||||||||||||||| ||||||||||| || ||
Db        387 GCTTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAGGCTCCTTACACGTCTCCC 446

Qy        301 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 360
              |||||||||||||||||||||||||||||||||| ||  | ||||| ||||| ||||||
Db        447 GATAAACTAAAGGATATATTTATGTTTATAACAAGGCAACTAAAGGGACTAGAAGATACA 506

Qy        361 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 420
              |||||||| || || |||||||||||||||||||||||||||||||| ||||| || |||
Db        507 AAGAGCCCTCAGTTTAATAGGTATTTTTATTTACTTGAGAACATTGCATGGGTTAAATCA 566

Qy        421 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 480
              |||||||||||||||||||||||||||||||||||||| || || || ||||||||
Db        567 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATCTTTACTCAACTTTACAGAACA 626

Qy        481 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 540
              ||||| |||||||||||||||||||||||||||||||| |||||||||||||| |||||
Db        627 TTATTCTCAGTTATAAACAATGGCCACAATCAGAAAGTTCATATGCACATGGTGGACCTC 686

Qy        541 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 600
              ||||| ||||| ||||||||||||||||| || ||||||||||| || ||||| ||||||
Db        687 ATGAGTTCTATCATTTGTGAAGGTGATACCGTATCTCAGGAGCTCTTAGATACAGTTTTA 746

Qy        601 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 660
              ||||||||||||||||| ||||||||| ||||||||||||||||||||||||||||||||
Db        747 GTAAATCTGGTACCTGCCCATAAGAACTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 806

Qy        661 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCACTTTTTTTAATCAGGTT 720
              || ||||||||||| |||||||||||||| |||||||||||| ||||||||| ||||||
Db        807 TTGCTGAAGAGGACTGCTCAAGCTATTGAACCATATATTACCAATTTTTTTAACCAGGTT 866

Qy        721 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 780
              ||||||||||||||||||||||||| |||||||| |||||||||||||||| || || |||
Db        867 CTGATGCTTGGGAAAACATCTATCAGTGATTTGTCTGAGCATGTCTTTGATTTGATCTTG 926

Qy        781 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 840
              || |||||||| ||||| || ||||||||||| |||||| | ||||||||||| ||||||
Db        927 GAACTCTACAACATTGACAGCCATTTGCTGCTTTCTGTTCTGCCCCAGCTTGAGTTTAAA 986

Qy        841 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 900
              ||||||||||||||||||| || ||||| || ||||||||||||||||||| ||||||||
Db        987 TTAAAGAGCAATGATAATGAAGAACGCCTCCAGGTTGTTAAACTACTGGCGAAAATGTTT 1046

Qy        901 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 960
              |||||||||||||||||||||||||||||| || || || |||||||||||| ||||||
Db       1047 GGGGCAAAGGATTCAGAATTGGCTTCTCAGAACAAACCCCTCTGGCAGTGCTACCTGGGC 1106

Qy        961 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1020
              |||||||||||||||||||||||||| |||||||||||||||||||||||| ||||| ||
Db       1107 AGGTTTAATGATATCCATGTACCAATACGCCTGGAATGTGTGAAATTTGCCAGCCACTGC 1166

Qy       1021 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1080
              ||||||||||||||||||||||||||| || ||||||||||||||||||||||||||||||
Db       1167 CTCATGAACCATCCTGATTTAGCAAAGGATTTAACAGAGTATCTTAAAGTGAGGTCACAT 1226

Qy       1081 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1140
              || |||||||||||||||||||||||||||||||||||| ||||| |||||||||||||||
```

```
Db       1227 GATCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCTATAGTAACAGCTGCTAAAAAG 1286

Qy       1141 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1200
Db       1287 GATATTCTTCTTGTCAATGATCACTTACTCAATTTTGTGAGAGAGCGAACATTGGACAAA 1346

Qy       1201 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1260
Db       1347 CGGTGGAGAGTACGTAAAGAAGCCATGATGGGACTGGCCCAGATTTATAAGAAATACTCT 1406

Qy       1261 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1320
Db       1407 TTACAGTCAGCAGCTGGCAAAGATGCTGCAAAACAGATATCCTGGGTCAAGGACAAGCTG 1466

Qy       1321 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1380
Db       1467 CTGCATATATACTATCAGAACAGCATCGACGACCGGCTACTTGTTGAGCGGATCTTTGCT 1526

Qy       1381 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1440
Db       1527 CAGTACATGGTCCCTCACAATTTGGAAACGACGGAACGGATGAAGTGCTTGTATTACTTG 1586

Qy       1441 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1500
Db       1587 TACGCTACACTGGACTTGAATGCTGTGAAAGCATTGAATGAAATGCGGAAATGTCAAAAT 1646

Qy       1501 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1560
Db       1647 CTACTCCGACATCAAGTAAAGGATTTACTTGATTTAATTAAACAACCAAAAACAGATGCC 1706

Qy       1561 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1620
Db       1707 AGTGTCAAGGCCATATTTTCTAAAGTGATGGTTATCACAAGAAATTTGCCAGATCCTGGT 1766

Qy       1621 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1680
Db       1767 AAGGCCCAAGACTTCATGAAGAAGTTCACACAAGTGTTGGAAGATGATGAGAAAATACGG 1826

Qy       1681 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1740
Db       1827 AAACAGCTGGAAGCCCTTGTTAGTCCCACATGTTCCTGCAAGCAGGCTGAAGGCTGTGTG 1886

Qy       1741 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1800
Db       1887 CGTGAAATAACTAAGAAGTTGGGCAACCCTAAGCAGCCTACAAACCCTTTCTTGGAAATG 1946

Qy       1801 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 1860
Db       1947 ATCAAGTTTCTCTTGGAGAGGATAGCTCCTGTACACATAGACACTGAATCCATCAGTGCT 2006

Qy       1861 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 1920
Db       2007 CTTATCAAGCAAGTGAATAAATCAATAGATGGAACAGCAGATGATGAGGATGAGGGTGTT 2066

Qy       1921 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 1980
Db       2067 CCAACTGATCAAGCTATCAGGGCAGGCCTTGAGCTGCTGAAGGTGCTGTCTTTTACGCAC 2126

Qy       1981 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2040
Db       2127 CCTATCTCCTTTCACTCTGCTGAGACGTTTGAGTCCCTTCTGGCTTGTCTGAAGATGGAT 2186

Qy       2041 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2100
Db       2187 GATGAGAAGGTGGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2246

Qy       2101 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2160
Db       2247 GAGGATTTCCCTCACATCAGATCAGCCTTGCTTCCTGTACTACATCACAAATCCAAAAAA 2306

Qy       2161 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2220
Db       2307 GGACCACCTCGGCAGGCCAAATACGCCATCCATTGTATTCATGCCATATTTTCTAGTAAA 2366

Qy       2221 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2280
Db       2367 GAAACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAAAGCCTAGATCCAAGCAACCTG 2426

Qy       2281 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2340
Db       2427 GAGCATCTTATAACCCCTCTGGTCACTATTGGTCATATTGCTCTTCTCGCACCTGATCAG 2486

Qy       2341 TTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2400
Db       2487 TTCGCTGCTCCTCTGAAGTCTTTGGTGGCAACTTTCATTGTGAAGGACCTCCTCATGAAT 2546

Qy       2401 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2460
Db       2547 GACCGGCTTCCAGGAAAAAAGACAACTAAGCTTTGGGTTCCAGATGAGGAAGTCTCACCT 2606

Qy       2461 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2520
Db       2607 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAG 2666

Qy       2521 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2580
Db       2667 AATAATCACAGTAAGTCAGGAACTTCCACCCTCAGATTACTAACAACGATACTGCATAGT 2726

Qy       2581 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2640
Db       2727 GATGGGGATTTGACAGAACAAGGGAAAATTAGTAAACCAGATATGTCACGCCTGAGACTT 2786

Qy       2641 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2700
Db       2787 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAGGAGCCCTGTTACCACGAGATCATTACA 2846

Qy       2701 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2760
Db       2847 CTGGAGCAGTACCAGCTGTGTGCATTAGCCATCAATGATGAGTGTTATCAAGTCAGGCAG 2906

Qy       2761 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2820
Db       2907 GTGTTCGCCCAGAAACTTCACAAAGGCCTTTCCCGCTTACGGCTTCCCCTTGAGTACATG 2966

Qy       2821 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 2880
Db       2967 GCCATCTGTGCTCTTTGTGCCAAAGACCCTGTGAAGGAGAGGCGAGCCCATGCTAGACAG 3026

Qy       2881 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 2940
Db       3027 TGCTTGGTGAAAAACACCACTGTGAGGAGGGAGTACCTGAAGCAGCATGCAAGTGTTAGT 3086

Qy       2941 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3000
Db       3087 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3146

Qy       3001 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3060
Db       3147 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTGAAAGAATGC 3206

Qy       3061 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3120
Db       3207 CTTTGGTTTGTTCTGGAGATATTGATGGCTAAAAATGAAAACAACAGCCATGCATTTATC 3266

Qy       3121 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3180
Db       3267 AGAAAAATGGTAGAAAATATTAAACAGACAAAAGATGCTCAAGGACCAGATGATACAAAA 3326

Qy       3181 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3240
Db       3327 ATGAATGAAAAATTGTACACCGTGTGTGATGTTGCCATGAACATCATCATGTCCAAGAGC 3386

Qy       3241 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3300
Db       3387 ACCACGTACAGCCTGGAGTCTCCTAAGGACCCCGTGCTGCCAGCTCGGTTTTTCACCCAG 3446
```

```
Qy       3301 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3360
Db       3447 CCTGACAAGAATTTTAGTAACACCAAAAATTACCTGCCTCCAGAAATGAAATCATTTTTC 3506

Qy       3361 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3420
Db       3507 ACTCCTGGAAAACCTAAAACAGCCAATGTTCTCGGAGCAGTTAATAAGCCACTTTCATCA 3566

Qy       3421 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3480
Db       3567 GCAGGCAAACAGTCTCAGACCAAATCATCAAGAATGGAAACTGTGAGCAACGCAAGCAGC 3626

Qy       3481 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3540
Db       3627 AGCTCCAACCCAAGCTCTCCTGGAAGGATCAAGGGGAGGCTTGATAGCTCTGAAATGGAT 3686

Qy       3541 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3600
Db       3687 CACAGTGAAAATGAAGATTATACAATGTCTTCACCTTTGCCAGGAAAAAAAAGTGACAAG 3746

Qy       3601 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3660
Db       3747 AGAGAAGACCCTGATCTT------TCTGAGTTGGAGAAGCCTAGAAGTCGGAAAAAAGCA 3800

Qy       3661 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3720
Db       3801 CCTGTCACAGACCCTGAAGAGAAATTAGGTATGGATGACCTAACTAAGTTGGTACAGGAA 3860

Qy       3721 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3780
Db       3861 CAGAAACCTAAAGGCAGTCAGCGAGGCCGGAAAAGAGGCCGTACAGCTTCAGACTCAGAC 3920

Qy       3781 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3840
Db       3921 GAGCAGCAGTGGCCTGAGGAGAAGAGGCACAAAGAGGAGCTCCTGGAAAATGAGGATGAG 3980

Qy       3841 CAGAATAGTCCGCCAAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 3900
Db       3981 CAGAACAGCCCACCAAAAAAAGGGCAAAAGAGGCAGGCCACCAAAACCTCTTGGTGGGGGG 4040

Qy       3901 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 3960
Db       4041 ACATCGAAGGAAGAGCCCACAATGAAAACATCCAAGAAAGGAAACAAGAAAAAACTTGTA 4100

Qy       3961 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4020
Db       4101 CCTCCTGTAGTAGACGACGATGAAGAAGAAGAAAGACAAATTGGAAACACAGAACATAAG 4160

Qy       4021 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4080
Db       4161 TCAAAAAGCAAACAGCACCGCACATCAAAGAGAGCACAACAGAGAGCAGAATCTCCTGAA 4220

Qy       4081 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4140
Db       4221 ACAAGTGCAGTTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4280

Qy       4141 ACGCCATCACCATCACAACCAAAAAAAAA 4169
Db       4281 GCACCATCACCATCACAACCCCCCAAAAA 4309

RESULT 5
AK122414
LOCUS       AK122414                5944 bp    mRNA    linear   ROD 15-MAR-2003
DEFINITION  Mus musculus mRNA for mKIAA0979 protein.
ACCESSION   AK122414
VERSION     AK122414.1  GI:28972559
KEYWORDS    FLI_CDNA.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Aizawa,H., Yuasa,S.,
            Nakajima,D., Nagase,T., Ohara,O. and Koga,H.
  TITLE     Prediction of the coding sequences of mouse homologues of KIAA
            gene: II. The complete nucleotide sequences of 400 mouse
            KIAA-homologous cDNAs identified by screening of terminal sequences
            of cDNA clones randomly sampled from size-fractionated libraries
  JOURNAL   DNA Res. 10, 35-48 (2003)
REFERENCE   2  (bases 1 to 5944)
  AUTHORS   Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research
            Institute, Laboratory for Genome Informatics; 2-6-7
            Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
            (E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
COMMENT     The CREATE program supported by Japan science and technology
            corporation; cDNA full insert sequencing: Kazusa DNA Research
            Institute; cDNA library construction, clone selection and 5'- &
            3'-end one pass sequencing.
FEATURES             Location/Qualifiers
     source          1..5944
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="mbx00002"
                     /tissue_type="brain"
                     /dev_stage="adult"
                     /note="vector:modified pBC SK+"
     gene            1..5944
                     /gene="mKIAA0979"
     CDS             <1..3577
                     /gene="mKIAA0979"
                     /note="CDS is predicted by in silico analysis. Start codon
                     is not identified."
                     /codon_start=2
                     /evidence=not_experimental
                     /product="mKIAA0979 protein"
                     /protein_id="BAC65696.1"
                     /db_xref="GI:28972560"
                     /translation="FDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKM
                     FGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKV
                     RSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQI
                     YKKYSLQSAAGKDAAKQISWVKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTER
                     MKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMV
                     ITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEALVSPTCSCKQAEGCVREITKKLGN
                     PKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIR
                     AGLELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPH
                     IRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHL
                     ITPLVTIGHIALLAPDQFAAPLKSLVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPE
                     TMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLR
                     LAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPL
                     EYMAICALCAKDPVKERRAHARQCLVKNITVRREYLKQHAAVSEKLLSLLPEYVVPYT
                     IHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDA
                     QGPDDTKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNY
                     LPPEMKSFFTPGKPKTANVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGR
                     IKGRLDSSEMDHSENEDYTMSSPLPGKKSDKREDPDLSELEKPRSRKKAPVTDPEEKL
                     GMDDLTKLVQEQKPKGSQRGRKRGRTASDSDEQQWPEEKRHKEELLENEDEQNSPPKK
                     GKRGRPPKPLGGGTSKEEPTMKTSKKGNKKKLVPPVVDDDEEEERQIGNTEHKSKSKQ
                     HRTSKRAQQRAESPETSAVESTQSTPQKGRGRPSKAPSPSQPPKKIRVGRSKQVATKE
                     NDSSEEMDVLQASSPVSDDTTQEGAEEEDISVGNVRRRSSKRERR"
BASE COUNT     1892 a   1151 c   1251 g   1650 t
ORIGIN

  Query Match             67.5%;  Score 2816.6;  DB 10;  Length 5944;
  Best Local Similarity   89.4%;  Pred. No. 0;
  Matches 3045;  Conservative    0;  Mismatches  354;  Indels    6;  Gaps    1;

Qy        765 CTTTGACTTAATTTTGGAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTACC 824
Db          1 CTTTGATTTGATCTTGGAACTCTACAACATTGACAGCCATTTGCTGCTTTCTGTTCTGCC 60

Qy        825 CCAGCTTGAATTTAAATTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACT 884
Db         61 CCAGCTTGAGTTTAAATTAAAGAGCAATGATAATGAAGAACGCCTCCAGGTTGTTAAACT 120
```

```
Qy    885 ACTGGCAAAAATGTTTGGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTG 944
Db    121 ACTGGCGAAAATGTTTGGGGCAAAGGATTCAGAATTGGCTTCTCAGAACAAACCCCTCTG 180

Qy    945 GCAGTGCTACTTGGGCAGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAA 1004
Db    181 GCAGTGCTACCTGGGCAGGTTTAATGATATCCATGTACCAATACGCCTGGAATGTGTGAA 240

Qy   1005 ATTTGCTAGCCATTGTCTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCT 1064
Db    241 ATTTGCCAGCCACTGCCTCATGAACCATCCTGATTTAGCAAAGGATTTAACAGAGTATCT 300

Qy   1065 TAAAGTGAGGTCACATGACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGT 1124
Db    301 TAAAGTGAGGTCACATGATCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCTATAGT 360

Qy   1125 TACAGCTGCTAAAAAGGATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGA 1184
Db    361 AACAGCTGCTAAAAAGGATATTCTTCTTGTCAATGATCACTTACTCAATTTTGTGAGAGA 420

Qy   1185 GAGAACATTAGACAAACGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAAT 1244
Db    421 GCGAACATTGGACAAACGGTGGAGAGTACGTAAAGAAGCCATGATGGGACTGGCCCAGAT 480

Qy   1245 TTATAAGAAATATGCTTTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATG 1304
Db    481 TTATAAGAAATACTCTTTACAGTCAGCAGCTGGCAAAGATGCTGCAAAACAGATATCCTG 540

Qy   1305 GATCAAAGACAAATTGCTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGT 1364
Db    541 GGTCAAGGACAAGCTGCTGCATATATACTATCAGAACAGCATCGACGACCGGCTACTTGT 600

Qy   1365 TGAACGGATCTTTGCTCAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAA 1424
Db    601 TGAGCGGATCTTTGCTCAGTACATGGTCCCTCACAATTTGGAAACGACGGAACGGATGAA 660

Qy   1425 ATGCTTATATTACTTGTATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAAT 1484
Db    661 GTGCTTGTATTACTTGTACGCTACACTGGACTTGAATGCTGTGAAAGCATTGAATGAAAT 720

Qy   1485 GTGGAAATGTCAAAATCTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCA 1544
Db    721 GTGGAAATGTCAAAATCTACTCCGACATCAAGTAAAGGATTTACTTGATTTAATTAAACA 780

Qy   1545 ACCCAAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAA 1604
Db    781 ACCAAAAACAGATGCCAGTGTCAAGGCCATATTTTCTAAAGTGATGGTTATCACAAGAAA 840

Qy   1605 TTTACCTGATCCTGGTAAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGA 1664
Db    841 TTTGCCAGATCCTGGTAAGGCCCAAGACTTCATGAAGAAGTTCACACAAGTGTTGGAAGA 900

Qy   1665 TGATGAGAAAATAAGAAAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCA 1724
Db    901 TGATGAGAAAATACGGAAACAGCTGGAAGCCCTTGTTAGTCCCACATGTTCCTGCAAGCA 960

Qy   1725 GGCTGAAGGTTGTGTGCGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAA 1784
Db    961 GGCTGAAGGCTGTGTGCGTGAAATAACTAAGAAGTTGGGCAACCCTAAGCAGCCTACAAA 1020

Qy   1785 TCCTTTCCTGGAAATGATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATAC 1844
Db   1021 CCCTTTCTTGGAAATGATCAAGTTTCTCTTGGAGAGGATAGCTCCTGTACACATAGACAC 1080

Qy   1845 CGAATCTATCAGTGCTCTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGA 1904
Db   1081 TGAATCCATCAGTGCTCTTATCAAGCAAGTGAATAAATCAATAGATGGAACAGCAGATGA 1140

Qy   1905 TGAAGATGAGGGTGTTCCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGT 1964
Db   1141 TGAGGATGAGGGTGTTCCAACTGATCAAGCTATCAGGGCAGGCCTTGAGCTGCTGAAGGT 1200

Qy   1965 ACTCTCATTTACACATCCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGC 2024
Db   1201 GCTGTCTTTTACGCACCCTATCTCCTTTCACTCTGCTGAGACGTTTGAGTCCCTTCTGGC 1260

Qy   2025 TTGTCTGAAAATGGATGATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACAC 2084
Db   1261 TTGTCTGAAGATGGATGATGAGAAGGTGGCAGAAGCTGCACTACAAATTTTCAAAAACAC 1320

Qy   2085 AGGAAGCAAAATTGAAGAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACA 2144
Db   1321 AGGAAGCAAAATTGAAGAGGATTTCCCTCACATCAGATCAGCCTTGCTTCCTGTACTACA 1380

Qy   2145 TCACAAATCTAAAAAAGGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGC 2204
Db   1381 TCACAAATCCAAAAAAGGACCACCTCGGCAGGCCAAATACGCCATCCATTGTATTCATGC 1440

Qy   2205 GATATTTTCTAGTAAAGAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCT 2264
Db   1441 CATATTTTCTAGTAAAGAAACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAAAGCCT 1500

Qy   2265 AGATCCAAGCAACCTGGAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCT 2324
Db   1501 AGATCCAAGCAACCTGGAGCATCTTATAACCCCTCTGGTCACTATTGGTCATATTGCTCT 1560

Qy   2325 CCTTGCACCTGATCAATTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAA 2384
Db   1561 TCTCGCACCTGATCAGTTCGCTGCTCCTCTGAAGTCTTTGGTGGCAACTTTCATTGTGAA 1620

Qy   2385 AGATCTTCTCATGAATGATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGA 2444
Db   1621 GGACCTCCTCATGAATGACCGGCTTCCAGGAAAAAAGACAACTAAGCTTTGGGTTCCAGA 1680

Qy   2445 TGAAGAAGTATCTCCTGAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATG 2504
Db   1681 TGAGGAAGTCTCACCTGAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATG 1740

Qy   2505 GCTACTTGGAATGAAAAATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAAC 2564
Db   1741 GCTACTTGGAATGAAGAATAATCACAGTAAGTCAGGAACTTCCACCCTCAGATTACTAAC 1800

Qy   2565 AACAATATTGCATAGTGATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATAT 2624
Db   1801 AACGATACTGCATAGTGATGGGGATTTGACAGAACAAGGGAAAATTAGTAAACCAGATAT 1860

Qy   2625 GTCACGTCTGAGACTTGCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTA 2684
Db   1861 GTCACGCCTGAGACTTGCTGCTGGGAGTGCTATTGTGAAGCTGGCACAGGAGCCCTGTTA 1920

Qy   2685 CCATGAAATCATCACATTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATG 2744
Db   1921 CCACGAGATCATTACACTGGAGCAGTACCAGCTGTGTGCATTAGCCATCAATGATGAGTG 1980

Qy   2745 CTATCAAGTAAGACAAGTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCT 2804
Db   1981 TTATCAAGTCAGGCAGGTGTTCGCTCAGAAACTTCACAAAGGCCTTTCCCGCTTACGGCT 2040

Qy   2805 TCCACTTGAGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAG 2864
Db   2041 TCCCCTTGAGTACATGGCCATCTGTGCTCTTTGTGCCAAAGACCCTGTGAAAGAGAGGCG 2100

Qy   2865 AGCTCATGCTAGGCAATGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCA 2924
Db   2101 AGCCCATGCTAGACAGTGCTTGGTGAAGAACATCACTGTGAGGAGGGAGTACCTGAAGCA 2160

Qy   2925 GCATGCAGCTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATAC 2984
Db   2161 GCATGCAGCTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATAC 2220

Qy   2985 AATTCACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAA 3044
Db   2221 AATTCACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAA 2280

Qy   3045 AGATGTTAAAGAATGTCTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAA 3104
```

```
Db       2281 AGATGTGAAAGAATGCCTTTGGTTTGTTCTGGAGATATTGATGGCTAAAAATGAAAACAA 2340

Qy       3105 CAGTCACGCTTTTATCAGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGG 3164
Db       2341 CAGCCATGCATTTATCAGAAAAATGGTAGAAAATATTAAACAGACAAAAGATGCTCAAGG 2400

Qy       3165 ACCAGATGATGCAAAAATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATAT 3224
Db       2401 ACCAGATGATACAAAAATGAATGAAAAATTGTACACCGTGTGTGATGTTGCCATGAACAT 2460

Qy       3225 CATCATGTCAAAGAGTACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGC 3284
Db       2461 CATCATGTCCAAGAGCACCACGTACAGCCTGGAGTCTCCTAAGGACCCCGTGCTGCCAGC 2520

Qy       3285 TCGTTTCTTCACTCAACCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGA 3344
Db       2521 TCGGTTTTTCACCCAGCCTGACAAGAATTTTAGTAACACCAAAAATTACCTGCCTCCAGA 2580

Qy       3345 AATGAAATCATTTTTCACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAA 3404
Db       2581 AATGAAATCATTTTTCACTCCTGGAAAACCTAAAACAGCCAATGTTCTCGGAGCAGTTAA 2640

Qy       3405 CAAGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGT 3464
Db       2641 TAAGCCACTTTCATCAGCAGGCAAACAGTCTCAGACCAAATCATCAAGAATGGAAACTGT 2700

Qy       3465 AAGCAATGCAAGCAGCAGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGA 3524
Db       2701 GAGCAACGCAAGCAGCAGCTCCAACCCAAGCTCTCCTGGAAGGATCAAGGGGAGGCTTGA 2760

Qy       3525 TAGTTCTGAAATGGATCACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGG 3584
Db       2761 TAGCTCTGAAATGGATCACAGTGAAAATGAAGATTATACAATGTCTTCACCTTTGCCAGG 2820

Qy       3585 GAAAAAAAGTGACAAGAGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAG 3644
Db       2821 AAAAAAAAGTGACAAGAGAGAAGACCCTGATCTT------TCTGAGTTGGAGAAGCCTAG 2874

Qy       3645 AGGCAGGAAAAAAACGCCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGAC 3704
Db       2875 AAGTCGGAAAAAAGCACCTGTCACAGACCCTGAAGAGAAATTAGGTATGGATGACCTAAC 2934

Qy       3705 TAAGTTGGTACAGGAACAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATAC 3764
Db       2935 TAAGTTGGTACAGGAACAGAAACCTAAAGGCAGTCAGCGAGGCCGGAAAAGAGGCCGTAC 2994

Qy       3765 GGCTTCAGAATCTGATGAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATT 3824
Db       2995 AGCTTCAGACTCAGACGAGCAGCAGTGGCCTGAGGAGAAGAGGCACAAAGAGGAGCTCCT 3054

Qy       3825 AGAAAATGAAGATGAACAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAA 3884
Db       3055 GGAAAATGAGGATGAGCAGAACAGCCCACCAAAAAAGGGCAAAAGAGGCAGGCCACCAAA 3114

Qy       3885 ACCTCTTGGTGGAGGTACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAG 3944
Db       3115 ACCTCTTGGTGGGGGACATCGAAGGAAGAGCCCACAATGAAAACATCCAAGAAAGGAAA 3174

Qy       3945 CAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGG 4004
Db       3175 CAAGAAAAAACTTGTACCTCCTGTAGTAGACGACGATGAAGAAGAAGAAAGACAAATTGG 3234

Qy       4005 AAATACGGAACAGAAGTCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAG 4064
Db       3235 AAACACAGAACATAAGTCAAAAAGCAAACAGCACCGCACATCAAAGAGAGCACAACAGAG 3294

Qy       4065 AGCAGAATCTCCTGAATCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACG 4124
Db       3295 AGCAGAATCTCCTGAAACAAGTGCAGTTGAATCCACACAGTCCACACCACAGAAAGGACG 3354

Qy       4125 AGGAAGACCATCAAAAACGCCATCACCATCACAACCAAAAAAAAA 4169
Db       3355 AGGAAGACCATCAAAAGCACCATCACCATCACAACCCCCCAAAAA 3399


RESULT 6
AF294791
LOCUS       AF294791                6744 bp    mRNA    linear   PRI 24-JUL-2002
DEFINITION  Homo sapiens SCC-112 (SCC-112) mRNA, complete cds.
ACCESSION   AF294791
VERSION     AF294791.1  GI:21951801
KEYWORDS    .
SOURCE      Homo sapiens (human) .
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 6744)
  AUTHORS   Kumar,D., Patel,S., Whiteside,T.L. and Kasid,U.
  TITLE     Identification and characterization of SCC-112, a novel cell cycle
            regulated gene in human cancer cells
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6744)
  AUTHORS   Kumar,D. and Kasid,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2000) Radiation, Medicine, 3970 Reservoir Rd, NW,
            Washington, DC 20007, USA
FEATURES             Location/Qualifiers
     source          1..6744
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
     gene            1..6744
                     /gene="SCC-112"
     CDS             232..4125
                     /gene="SCC-112"
                     /codon_start=1
                     /product="SCC-112"
                     /protein_id="AAM82347.1"
                     /db_xref="GI:21951802"
                     /translation="MIKRLKMVVKTFMDMDQDSEDEKQQYLPLALHLASEFFLRNPNK
                     DVRLLVACCLADIFRIYAPEAPYTSHDKLKDIFLFITRQLKGLEDTKSPQFNRYFYLL
                     ENLAWVKSYNICFELEDCNEIFIQLFRTLFSVINNSHNKKVQMHMLDLMSSIIMEGDG
                     VTQELLGSILINLIPAHKNLNKQSFDLAKVLLKRTVQTIEACIANFFNQVLVLGRSSV
                     SDLSEHVFDLIQELFAIDPHLLLSVMPQLEFKLKSNDGEERLAVVRLLAKLFGSKDSD
                     LATQNRPLWQCFLGRFNDIHVPVRLESVKFASHCLMNHPDLAKDLTEYLKVRSHDPEE
                     AIRHDVIVTIITAAKRDLALVNDQLLGFVRERTLDKRWRVRKEAMMGLAQLYKKYCLH
                     GEAGKEAAEKVSWIKDKLLHIYYQNSIDDKLLVEKIFAQYLVPHNLETEERMKCLYYL
                     YASLDPNAVKALNEMWKCQNMLRSHVRELLDLHKQPTSEANCSAMFGKLMTIAKNLPD
                     PGKAQDFVKKFNQVLGDDEKLRSQLELLISPTCSCKQADICVREIARKLANPKQPTNP
                     FLEMVKFLLERIAPVHIDSEAISALVKLMNKSIEGTADDEEEGVSPDTAIRSGLELLK
                     VLSFTHPTSFHSAETYESLLQCLRMEDDKVAEAAIQIFRNTGHKIETDLPQIRSTLIP
                     ILHQKAKRGTPHQAKQAVHCIHAIFTNKEVQLAQIFEPLSRSLNADVPEQLITPLVSL
                     GHISMLAPDQFASPMKSVVANFIVKDLLMNDRSTGEKNGKLWSPDEEVSPEVLAKVQA
                     IKLLVRWLLGMKNNQSKSANSTLRLLSAMLVSEGDLTEQKRISKSDMSRLRLAAGSAI
                     MKLAQEPCYHEIITPEQFQLCALVINDECYQVRQIFAQKLHKALVKLLLPLEYMAIFA
                     LCAKDPVKERRAHARQCLLKNISIRREYIKQNPMATEKLLSLLPEYVVPYMIHLLAHD
                     PDFTRSQDVDQLRDIKECLWFMLEVLMTKNENNSHAFMKKMAENIKLTRDAQSPDESK
                     TNEKLYTVCDVALCVINSKSALCNADSPKDPVLPMKFFTQPEKDFCNDKSYISEETRV
                     LLLTGKPKPAGVLGAVNKPLSATGRKPYVRSTGTETGSNINVNSELNPSTGNRSREQS
                     SEAAETGVSENEENPVRIISVTPVKNIDPVKNKEINSDQATQGNISSDRGKKRTVTAA
                     GAENIQQKTDEKVDESGPPAPSKPRRGRRPKSESQGNATKNDDLNKPINKGRKRAAVG
                     QESPGGLEAGNAKAPKLQDLAKKAAPAERQIDLQR"
BASE COUNT     2156 a   1243 c   1400 g   1945 t
ORIGIN

  Query Match             43.5%;  Score 1815.6;  DB 9;  Length 6744;
  Best Local Similarity   69.7%;  Pred. No. 0;
  Matches 2457;  Conservative    0;  Mismatches 1069;  Indels    0;  Gaps    0;

Qy         32 GAAAAATTACATATCCGCCTGGGGTCAAGGAAATATCAGATAAAATATCTAAAGAGGAGA 91
Db        173 GGAAGATCGCTTACCCTCCGGGGGTAAAGAGATCACCGACAAGATCACCACGGACGAGA 232

Qy         92 TGGTGAGACGATTAAAGATGGTTGTGAAAACTTTTATGGATATGGACCAGGACTCTGAAG 151
```

```
Db        233 TGATCAAACGCCTGAAGATGGTAGTGAAAACCTTTATGGATATGGATCAGGACTCAGAAG 292

Qy        152 AAGAAAAGGAGCTTTATTTAAACCTAGCTTTACATCTTGCTTCAGATTTTTTCTCAAGC 211
Db        293 ATGAAAAACAGCAGTATCTCCCACTAGCCTTGCATCTTGCATCTGAATTCTTCCTCAGGA 352

Qy        212 ATCCTGGTAAAGATGTTCGCTTACTGGTAGCCTGCTGCCTTGCTGATATTTTCAGGATTT 271
Db        353 ACCCCAATAAAGATGTGCGTCTCCTTGTAGCATGTTGTTTGGCTGATATCTTTCGTATCT 412

Qy        272 ATGCTCCTGAAGCTCCTTACACATCCCCTGATAAACTAAAGGATATATTTATGTTTATAA 331
Db        413 ATGCCCCAGAAGCTCCATATACTTCCCATGATAAACTTAAGGACATATTTTTGTTTATTA 472

Qy        332 CAAGACAGTTGAAGGGGCTAGAGGATACAAAGAGCCCACAATTCAATAGGTATTTTTATT 391
Db        473 CCAGACAATTAAAAGGTTTGGAGGATACAAAGAGTCCACAGTTTAATAGATACTTTTATT 532

Qy        392 TACTTGAGAACATTGCTTGGGTCAAGTCATATAACATATGCTTTGAGTTAGAAGATAGCA 451
Db        533 TATTAGAGAATTTAGCTTGGGTTAAATCATATAACATCTGCTTTGAATTGGAAGATTGCA 592

Qy        452 ATGAAATTTTCACCCAGCTATACAGAACCTTATTTTCAGTTATAAACAATGGCCACAATC 511
Db        593 ATGAAATTTTTATTCAGCTTTTTAGAACTCTCTTCTCAGTGATCAACAATAGCCACAATA 652

Qy        512 AGAAAGTCCATATGCACATGGTAGACCTTATGAGCTCTATTATTTGTGAAGGTGATACAG 571
Db        653 AGAAGGTACAAATGCACATGCTAGATTTGATGAGTTCTATCATCATGGAAGGTGATGGAG 712

Qy        572 TGTCTCAGGAGCTTTTGGATACGGTTTTAGTAAATCTGGTACCTGCTCATAAGAATTTAA 631
Db        713 TTACTCAAGAATTATTGGGCTCCATTCTTATTAACCTCATTCCTGCACATAAGAACTTAA 772

Qy        632 ACAAGCAAGCATATGATTTGGCAAAGGCTTTACTGAAGAGGACAGCTCAAGCTATTGAGC 691
Db        773 ATAAACAGTCCTTTGACCTTGCAAAAGTGCTATTGAAAAGAACAGTCCAGACTATTGAGG 832

Qy        692 CATATATTACCACTTTTTTTAATCAGGTTCTGATGCTTGGGAAAACATCTATCAGCGATT 751
Db        833 CATGCATTGCCAATTTTTTCAATCAAGTCCTGGTGCTGGGAAGATCATCAGTAAGTGATT 892

Qy        752 TGTCAGAGCATGTCTTTGACTTAATTTTGGAGCTCTACAATATTGATAGTCATTTGCTGC 811
Db        893 TGTCAGAACATGTATTTGATCTGATTCAGGAACTTTTTGCTATAGATCCTCATTTATTAT 952

Qy        812 TCTCTGTTTTACCCCAGCTTGAATTTAAATTAAAGAGCAATGATAATGAGGAGCGCCTAC 871
Db        953 TATCCGTCATGCCACAGCTTGAATTCAAACTAAAGAGCAATGATGGAGAAGAGCGATTAG 1012

Qy        872 AAGTTGTTAAACTACTGGCAAAAATGTTTGGGGCAAAGGATTCAGAATTGGCTTCTCAAA 931
Db       1013 CTGTTGTTCGACTTCTAGCTAAATTGTTTGGCTCCAAAGATTCTGATTTGGCAACACAGA 1072

Qy        932 ACAAGCCACTTTGGCAGTGCTACTTGGGCAGGTTTAATGATATCCATGTACCAATCCGCC 991
Db       1073 ATCGTCCTCTTTGGCAATGTTTTCTTGGACGATTTAATGATATTCATGTTCCTGTGAGAT 1132

Qy        992 TGGAATGTGTGAAATTTGCTAGCCATTGTCTCATGAACCATCCTGATTTAGCAAAAGACT 1051
Db       1133 TAGAAAGTGTGAAATTTGCCAGTCATTGTTTAATGAATCACCCAGATTTAGCGAAGGATC 1192

Qy       1052 TAACAGAGTATCTTAAAGTGAGGTCACATGACCCTGAGGAAGCTATTAGACATGATGTTA 1111
Db       1193 TCACAGAATATTTAAAGGTTAGATCACATGATCCAGAAGAAGCTATTCGTCATGATGTCA 1252

Qy       1112 TTGTGTCAATAGTTACAGCTGCTAAAAAGGATATTCTTCTGGTCAATGATCACTTACTTA 1171
Db       1253 TTGTTACTATAATAACAGCTGCCAAGAGGGACCTGGCCTTAGTAAATGATCAGCTGCTTG 1312

Qy       1172 ATTTTGTGAGAGAGAGAACATTAGACAAACGATGGAGAGTACGCAAAGAAGCCATGATGG 1231
Db       1313 GCTTTGTAAGGGAAAGAACACTGGATAAACGGTGGCGAGTAAGAAAAGAAGCTATGATGG 1372

Qy       1232 GACTTGCCCAAATTTATAAGAAATATGCTTTACAGTCAGCAGCTGGAAAAGATGCTGCAA 1291
Db       1373 GTCTGGCTCAGCTTTATAAGAAATACTGTCTTCATGGTGAAGCAGGAAAGGAAGCTGCAG 1432

Qy       1292 AACAGATAGCATGGATCAAAGACAAATTGCTACATATATATTATCAAAATAGTATTGATG 1351
Db       1433 AGAAAGTCAGCTGGATAAAGGACAAACTTCTGCATATTTATTATCAGAACAGCATTGACG 1492

Qy       1352 ATCGACTACTTGTTGAACGGATCTTTGCTCAATACATGGTTCCTCACAATTTAGAAACTA 1411
Db       1493 ACAAACTGTTGGTAGAGAAAATCTTTGCTCAGTATCTTGTCCCCCACAACCTGGAAACAG 1552

Qy       1412 CAGAACGGATGAAATGCTTATATTACTTGTATGCCACACTGGATTTAAATGCTGTGAAAG 1471
Db       1553 AAGAGAGAATGAAATGCTTATATTACTTATATGCTAGTTTGGATCCAAATGCTGTAAAAG 1612

Qy       1472 CATTGAATGAAATGTGGAAATGTCAAAATCTGCTCCGACATCAAGTAAAGGATTTGCTTG 1531
Db       1613 CTCTCAACGAAATGTGGAAGTGTCAGAACATGCTTCGGAGCCATGTACGCGAACTATTGG 1672

Qy       1532 ACTTGATTAAGCAACCCAAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGG 1591
Db       1673 ATTTGCACAAGCAGCCTACATCAGAGGCTAACTGTTCTGCCATGTTTGGAAAACTGATGA 1732

Qy       1592 TTATTACAAGAAATTTACCTGATCCTGGTAAGGCTCAGGATTTCATGAAGAAATTCACAC 1651
Db       1733 CCATAGCAAAGAATTTGCCTGACCCCGGGAAAGCACAAGATTTTGTGAAGAAATTTAACC 1792

Qy       1652 AGGTGTTAGAAGATGATGAGAAAATAAGAAAGCAGTTAGAAGTACTTGTTAGTCCAACAT 1711
Db       1793 AGGTTCTCGGCGATGATGAGAAACTTCGGTCTCAGTTGGAGTTATTAATTAGCCCAACCT 1852

Qy       1712 GCTCCTGCAAGCAGGCTGAAGGTTGTGTGCGTGAAATAACTAAGAAGTTGGGCAACCCCA 1771
Db       1853 GTTCTTGCAAACAAGCAGATATTTGTGTGAGAGAAATAGCCCGGAAACTTGCAAATCCTA 1912

Qy       1772 AACAGCCTACAAATCCTTTCCTGGAAATGATCAAGTTTCTCTTGGAGAGGATAGCACCTG 1831
Db       1913 AGCAACCAACAAATCCTTTTCTAGAGATGGTCAAATTTCTGTTGGAAAGAATCGCACCTG 1972

Qy       1832 TGCACATAGATACCGAATCTATCAGTGCTCTTATTAAACAAGTGAACAAATCAATAGATG 1891
Db       1973 TGCACATTGATTCAGAAGCCATAAGTGCACTAGTGAAATTGATGAATAAGTCAATAGAGG 2032

Qy       1892 GAACAGCAGATGATGAAGATGAGGGTGTTCCAACTGATCAAGCCATCAGAGCAGGTCTTG 1951
Db       2033 GGACAGCAGATGATGAAGAGGAGGGTGTAAGTCCAGATACAGCTATCCGTTCAGGACTTG 2092

Qy       1952 AACTGCTTAAGGTACTCTCATTTACACATCCCATCTCATTTCATTCTGCTGAAACATTTG 2011
Db       2093 AACTTCTTAAGGTTCTGTCTTTTACACATCCTACCTCGTTCCACTCTGCAGAGACATATG 2152

Qy       2012 AATCATTACTGGCTTGTCTGAAAATGGATGATGAAAAAGTAGCAGAAGCTGCACTACAAA 2071
Db       2153 AGTCCTTGTTACAGTGCCTAAGAATGGAGGATGACAAGGTAGCAGAAGCTGCTATTCAAA 2212

Qy       2072 TTTTCAAAAACACAGGAAGCAAAATTGAAGAGGATTTTCCACACATCAGATCAGCCTTGC 2131
Db       2213 TTTTTAGAAATACAGGTCACAAAATAGAAACAGACCTTCCCCAGATACGATCGACCTTAA 2272

Qy       2132 TTCCTGTTTTACATCACAAATCTAAAAAAGGACCCCCCCGTCAAGCCAAATATGCCATTC 2191
Db       2273 TTCCCATTTTACATCAAAAAGCAAAGAGGGGTACTCCACACCAAGCAAAACAGGCTGTGC 2332

Qy       2192 ATTGTATCCATGCGATATTTTCTAGTAAAGAGACCCAGTTTGCACAGATATTTGAGCCTC 2251
Db       2333 ACTGTATACACGCCATATTCACAAATAAAGAAGTCCAGCTTGCACAGATTTTTGAGCCAC 2392

Qy       2252 TGCATAAGAGCCTAGATCCAAGCAACCTGGAACATCTCATAACACCATTGGTTACTATTG 2311
Db       2393 TCAGTAGGAGTCTGAATGCTGATGTGCCAGAACAACTTATAACTCCATTAGTTTCATTGG 2452
```

```
Qy       2312 GTCATATTGCTCTCCTTGCACCTGATCAATTTGCTGCTCCTTGGAAATCTTGGGTAGCTA 2371
Db       2453 GCCACATTTCTATGTTAGCACCAGATCAGTTTGCTTCCCCAATGAAATCTGTAGTAGCAA 2512

Qy       2372 CTTTCATTGTGAAAGATCTTCTCATGAATGATCGGCTTCCAGGGAAAAAGACAACTAAAC 2431
Db       2513 ATTTTATTGTGAAAGATCTGCTAATGAATGACAGGTCAACAGGTGAAAAGAATGGAAAAC 2572

Qy       2432 TTTGGGTTCCAGATGAAGAAGTATCTCCTGAGACAATGGTCAAAATTCAGGCTATTAAAA 2491
Db       2573 TGTGGTCTCCAGATGAAGAGGTTTCCCCTGAAGTACTAGCAAAGGTACAGGCAATTAAAC 2632

Qy       2492 TGATGGTTCGATGGCTACTTGGAATGAAAAATAATCACAGTAAATCAGGAACTTCTACCT 2551
Db       2633 TTCTGGTAAGGTGGCTGTTGGGTATGAAAAACAACCAGTCTAAATCTGCCAATTCAACCC 2692

Qy       2552 TAAGATTGCTAACAACAATATTGCATAGTGATGGAGACTTGACAGAACAGGGGAAAATTA 2611
Db       2693 TTCGGTTATTATCAGCGATGTTGGTTAGTGAGGGTGACCTGACAGAGCAAAAGAGGATCA 2752

Qy       2612 GTAAACCAGATATGTCACGTCTGAGACTTGCTGCTGGGAGTGCTATTGTGAAGCTGGCAC 2671
Db       2753 GTAAATCTGATATGTCTCGCTTGCGATTAGCTGCTGGTAGTGCCATAATGAAGCTTGCTC 2812

Qy       2672 AAGAACCCTGTTACCATGAAATCATCACATTAGAACAATATCAGCTATGTGCATTAGCTA 2731
Db       2813 AGGAACCTTGTTACCATGAAATTATTACCCCAGAACAGTTTCAGCTCTGTGCACTTGTTA 2872

Qy       2732 TCAACGATGAATGCTATCAAGTAAGACAAGTGTTTGCCCAGAAACTTCACAAAGGCCTTT 2791
Db       2873 TTAATGATGAGTGTTACCAAGTAAGGCAGATATTTGCTCAGAAGCTGCATAAGGCACTTG 2932

Qy       2792 CCCGTTTACGGCTTCCACTTGAGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTG 2851
Db       2933 TGAAGTTACTGCTCCCATTGGAGTATATGGCGATCTTTGCCTTGTGTGCCAAAGATCCTG 2992

Qy       2852 TAAAGGAGAGAAGAGCTCATGCTAGGCAATGTTTGGTGAAAAATATAAATGTAAGGCGGG 2911
Db       2993 TGAAGGAGAGAAGAGCACACGCACGACAATGTTTACTGAAAAATATCAGTATACGCAGGG 3052

Qy       2912 AGTATCTGAAGCAGCATGCAGCTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATG 2971
Db       3053 AATACATTAAGCAGAATCCTATGGCTACTGAGAAATTATTATCACTGTTGCCTGAATATG 3112

Qy       2972 TTGTTCCATATACAATTCACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATA 3031
Db       3113 TAGTTCCATACATGATTCACCTGCTAGCCCATGATCCAGATTTTACAAGATCACAAGATG 3172

Qy       3032 TTGAACAACTTAAAGATGTTAAAGAATGTCTTTGGTTTGTTCTGGAAATATTAATGGCTA 3091
Db       3173 TTGATCAGCTTCGTGATATCAAAGAGTGCCTATGGTTCATGCTTGAAGTTTTAATGACAA 3232

Qy       3092 AAAATGAAAATAACAGTCACGCTTTTATCAGAAAGATGGTAGAAAATATTAAACAAACAA 3151
Db       3233 AGAATGAAAACAATAGCCATGCCTTTATGAAGAAGATGGCAGAGAACATCAAGTTAACCA 3292

Qy       3152 AAGATGCCCAAGGACCAGATGATGCAAAAATGAATGAAAAACTGTACACTGTGTGTGATG 3211
Db       3293 GAGATGCCCAGTCTCCAGATGAATCCAAGACAAATGAAAAACTGTATACAGTATGTGATG 3352

Qy       3212 TTGCCATGAATATCATCATGTCAAAGAGTACTACATACAGTTTGGAATCTCCTAAAGACC 3271
Db       3353 TGGCTCTCTGTGTTATAAATAGTAAAAGTGCTTTGTGCAATGCAGATTCACCAAAGGACC 3412

Qy       3272 CGGTACTACCAGCTCGTTTCTTCACTCAACCTGACAAGAATTTCAGTAACACCAAAAATT 3331
Db       3413 CAGTCCTCCCAATGAAATTTTTACACAACCTGAAAAGGACTTCTGTAACGATAAGAGTT 3472

Qy       3332 ATCTGCCTCCTGAAATGAAATCATTTTTCACTCCTGGAAAACCTAAAACAACCAATGTTC 3391
Db       3473 ATATTTCAGAAGAGACAAGAGTACTTCTGTTAACAGGAAAGCCAAAGCCTGCTGGAGTAC 3532

Qy       3392 TAGGAGCTGTTAACAAGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAAATCATCAC 3451
Db       3533 TAGGTGCAGTAAATAAGCCTTTATCAGCAACGGGAAGGAAACCCTATGTTAGAAGCACTG 3592

Qy       3452 GAATGGAAACTGTAAGCAATGCAAGCAGCAGCTCAAATCCAAGCTCTCCTGGAAGAATAA 3511
Db       3593 GCACTGAGACTGGAAGCAATATTAATGTAAATTCAGAGCTGAACCCTTCAACCGGAAATC 3652

Qy       3512 AGGGGAGGCTTGATAGTTCTGAAATGGATCACAGTGAAAATGAAGA 3557
Db       3653 GATCAAGGGAACAGAGTTCAGAGGCAGCAGAAACTGGAGTTAGTGA 3698
```

```
RESULT 7
HSU50533
LOCUS       HSU50533                1852 bp    mRNA    linear   PRI 27-NOV-1996
DEFINITION  Human BRCA2 region, mRNA sequence CG008.
ACCESSION   U50533
VERSION     U50533.1  GI:1531605
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1852)
  AUTHORS   Couch,F.J., Rommens,J.M., Neuhausen,S.L., Belanger,C., Dumont,M.,
            Kenneth,A., Bell,R., Berry,S., Bogden,R., Cannon-Albright,L.,
            Farid,L., Frye,C., Hattier,T., Janecki,T., Jiang,P., Kehrer,R.,
            Leblanc,J.-F., McArthur-Morrison,J., McSweeney,D., Miki,Y.,
            Peng,Y., Samson,C., Schroeder,M., Snyder,S.C., Stringfellow,M.,
            Stroup,C., Swedlund,B., Swensen,J., Teng,D., Thakur,S., Tran,T.,
            Tranchant,M., Welver-Feldhaus,J., Wong,A.K.C., Shizuya,H.,
            Labrie,F., Skolnick,M.H., Goldgar,D.E., Kamb,A., Weber,B.L.,
            Tavtigian,S.V. and Simard,J.
  TITLE     Generation of an integrated transcription map of the BRCA2 region
            on chromosome 13q12-q13
  JOURNAL   Genomics 36 (1), 86-99 (1996)
  MEDLINE   96411650
   PUBMED   8812419
REFERENCE   2  (bases 1 to 1852)
  AUTHORS   Simard,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-1996) Jacques Simard, Laboratory of Molecular
            Endocrinology, CHUL Research Center, 2705, Boulevard Laurier,
            Quebec City, Quebec G1V 4G2, Canada
FEATURES             Location/Qualifiers
     source          1..1852
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /map="13q12-q13"
                     /note="CG008; DSEG numbers: D13S171 and D13S310"
BASE COUNT      609 a    348 c    379 g    505 t     11 others
ORIGIN

  Query Match             42.0%;  Score 1754.4;  DB 9;  Length 1852;
  Best Local Similarity   98.0%;  Pred. No. 0;
  Matches 1817;  Conservative   10;  Mismatches   21;  Indels    6;  Gaps    5;

Qy       1064 TTAAAGTGAGGTCACATGACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAG 1123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTAAAGTGAGGTCACATGACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAG 60

Qy       1124 TTACAGCTGCTAAAAAGGATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAG 1183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTACAGCTGCTAAAAAGGATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAG 120

Qy       1184 AGAGAACATTAGACAAACGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAA 1243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGAGAACATTAGACAAACGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAA 180

Qy       1244 TTTATAAGAAATATGCTTTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCAT 1303
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTATAAGAAATATGCTTTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCAT 240

Qy       1304 GGATCAAAGACAAATTGCTACATATATATTATCAAAATAGTATTGATGATCGACTACTTG 1363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGATCAAAGACAAATTGCTACATATATATTATCAAAATAGTATTGATGATCGACTACTTG 300

Qy       1364 TTGAACGGATCTTTGCTCAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGA 1423
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTGAACGGATCTTTGCTCAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGA 360

Qy       1424 AATGCTTATATTACTTGTATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAA 1483
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AATGCTTATATTACTTGTATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAA 420

Qy       1484 TGTGGAAATGTCAAAATCTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGC 1543
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGTGGAAATGTCAAAATCTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGC 480

Qy       1544 AACCCAAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAA 1603
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AACCCAAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAA 540

Qy       1604 ATTTACCTGATCCTGGTAAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAG 1663
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATTTACCTGATCCTGGTAAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAG 600

Qy       1664 ATGATGAGAAAATAAGAAAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGC 1723
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATGATGAGAAAATAAGAAAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGC 660

Qy       1724 AGGCTGAAGGTTGTGTGCGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAA 1783
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGGCTGAAGGTTGTGTGCGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAA 720

Qy       1784 ATCCTTTCCTGGAAATGATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATA 1843
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATCCTTTCCTGGAAATGATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATA 780

Qy       1844 CCGAATCTATCAGTGCTCTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATG 1903
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CCGAATCTATCAGTGCTCTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATG 840

Qy       1904 ATGAAGATGAGGGTGTTCCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGG 1963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATGAAGATGAGGGTGTTCCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGG 900

Qy       1964 TACTCTCATTTACACATCCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGG 2023
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TACTCTCATTTACACATCCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGG 960

Qy       2024 CTTGTCTGAAAATGGATGATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACA 2083
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTTGTCTGAAAATGGATGATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACA 1020

Qy       2084 CAGGAAGCAAAATTGAAGAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTAC 2143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGGAAGCAAAATTGAAGAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTAC 1080

Qy       2144 ATCACAAATCTAAAAAAGGACCCCCCCG-TCAAGCCAAATATGCCATTCATTGTATCCAT 2202
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       1081 ATCACAAATCTAAAAAAGGACCCCCCCGCTCAAGCCAAATATGCCATTCATTGTATCCAT 1140

Qy       2203 GCGATATTTTCTAGTAAAGAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGC 2262
              |||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db       1141 GCGATATTTTCTAGTAAAGAGACYCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGC 1200

Qy       2263 CTAGATCCAAGCAACCTGGAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCT 2322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CTAGATCCAAGCAACCTGGAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCT 1260

Qy       2323 CTCCTTGCACCTGATCAATTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTG 2382
              ||||||||||||||||||||||||||||||| |||||||| |||||||||||||||||:|
Db       1261 CTCCTTGCACCTGATCAATTTGCTGCTCCTTTGAAATCTTTGGTAGCTACTTTCATTGYG 1320

Qy       2383 AAAGATCTTCTCATGAATGATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCA 2442
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAAGATCTTCTCATGAATGATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCA 1380

Qy       2443 GATGAAGAAGTATCTCCTGAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGA 2502
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GATGAAGAAGTATCTCCTGAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGA 1440

Qy       2503 TGGCTACTTGGAATGAAAAATAATCACAGTAAATCAGGAACTTCTACCTTAAG-ATTGCT 2561
              ||||||||||| ||| ||: ||:|||||||| ||||||| ||||| ||||||| ||||||
Db       1441 TGGCTACTTGGCATGTAAMTTAMTCACAGTATATCAGGATCTTCTTCCTTAAGSATTGCT 1500

Qy       2562 AACAACAATATTGCATAGTGATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGA 2621
              ||||||||||| |||| ||||||| || ||| :|||:|||||||||  |    |||||||
Db       1501 AACAACAATATCGCATCGTGATGGGGA-TTGGMAGAMCAGGGGAAA--TTAGTAACCAGA 1557

Qy       2622 TATGTCACGTCTGAGACTTGCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTG 2681
              |||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db       1558 TATGTCACGTCTGAGACTTGCTGCTGGGAGTGCTATTRTGAAGCTGGCACAAGAACCCTG 1617

Qy       2682 TTACCATGAAATCATCACATTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGA 2741
              |||||||||||||:|||||:||||||||||||||||||||||:|||||||||||||||||
Db       1618 TTACCATGAAATCRTCACAWTAGAACAATATCAGCTATGTGCRTTAGCTATCAACGATGA 1677

Qy       2742 ATGCTATCAAGTAAGACAAGTGTTT-GCCCAGAAACTTCACAAAGGCCTTTCCCGTTTAC 2800
              ||||||||||||||||||||||||| ||||||||||||| |||||||||||||||||| |
Db       1678 ATGCTATCAAGTAAGACAAGTGTTTGGCCCAGAAACTTCGCAAAGGCCTTTCCCGTTTGC 1737

Qy       2801 GGCTTCCACTTGAGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGA 2860
              |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db       1738 GGCTTCCACTTGAGAATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGA 1797

Qy       2861 GAAGAGCTCATGCTAGGCAATGTTTGGTGAAAAATATAAATGTAAGGCGGGAGT 2914
              ||||||||||||||||||||||||||||||||||||||||||||| |||||| |
Db       1798 GAAGAGCTCATGCTAGGCAATGTTTGGTGAAAAATATAAATGTAAAGCGGGACT 1851
```

```
RESULT 8
BC039256
LOCUS       BC039256                4252 bp    mRNA    linear   PRI 06-NOV-2002
DEFINITION  Homo sapiens, Similar to androgen-induced prostate proliferative
            shutoff associated protein, clone MGC:33757 IMAGE:5295109, mRNA,
            complete cds.
ACCESSION   BC039256
VERSION     BC039256.1  GI:24657778
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4252)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:      http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.
```

```
                 Clone distribution: MGC clone distribution information can be found
                 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
                 Series: IRAK Plate: 48 Row: k Column: 2
                 This clone was selected for full length sequencing because it
                 passed the following selection criteria: matched mRNA gi: 7705287.
FEATURES             Location/Qualifiers
     source          1. .4252
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:33757 IMAGE:5295109"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
     CDS             71. .1660
                     /codon_start=1
                     /product="Similar to androgen-induced prostate
                     proliferative shutoff associated protein"
                     /protein_id="AAH39256.1"
                     /db_xref="GI:24657779"
                     /translation="MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMD
                     MDQDSEEEKELYLNLALHLASDFFLKHPDKDVRLLVACCLADIFRIYAPEAPYTSPDK
                     LKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRT
                     LFSVINNGHNQKVHMHMVDLMSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLA
                     KALLKRTAQAIEPYITNFFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQ
                     LEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDINVPIRLECV
                     KFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNF
                     VRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSID
                     DRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKYVSNIKFCSFHPLQYIGF
                     YGKETTNTCILKCNLCSVNIV"
BASE COUNT     1407 a    664 c    792 g   1389 t
ORIGIN

  Query Match             35.1%;  Score 1464.6;  DB 9;  Length 4252;
  Best Local Similarity   95.5%;  Pred. No. 2.3e-291;
  Matches 1518;  Conservative    0;  Mismatches   69;  Indels    2;  Gaps    1;

Qy          1 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 130

Qy         61 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 190

Qy        121 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 250

Qy        181 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGGTAAAGATGTTCGCTTACTGGTA 240
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        251 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGATAAAGATGTTCGCTTACTGGTA 310

Qy        241 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 370

Qy        301 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        371 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 430

Qy        361 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        431 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 490

Qy        421 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        491 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 550

Qy        481 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        551 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 610

Qy        541 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        611 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 670

Qy        601 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        671 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 730

Qy        661 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCACTTTTTTTAATCAGGTT 720
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        731 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTTTTAATCAGGTT 790

Qy        721 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        791 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 850

Qy        781 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 910

Qy        841 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 970

Qy        901 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        971 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 1030

Qy        961 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1020
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1031 AGGTTTAATGATATCAATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1090

Qy       1021 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1091 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1150

Qy       1081 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1210

Qy       1141 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1270

Qy       1201 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1271 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1330

Qy       1261 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1331 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1390

Qy       1321 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1391 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1450

Qy       1381 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1451 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1510

Qy       1441 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1500
              |||||||||||||||||||||||||||||     ||  | |||| |  || | |     |
Db       1511 TATGCCACACTGGATTTAAATGCTGTGAA--GTATGTTTCAAATATCAAATTCTGTTCTT 1568

Qy       1501 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1560
               |  |||      |  ||  ||||||  | ||   ||     ||| |      |  |
Db       1569 TTCATCCTTTGCAATATATTGGATTTTATGGAAAAGAAACTACAAACACCTGTATTTTAA 1628

Qy       1561 AGTGTCAAGGCCATATTTTCAAAAGTGAT 1589
              | ||| |      || | | || | || |
Db       1629 AATGTAATCTATGTAGTGTAAACATTGTT 1657
```

```
RESULT 9
AB014548
LOCUS       AB014548                5177 bp    mRNA    linear   PRI 06-FEB-1999
DEFINITION  Homo sapiens mRNA for KIAA0648 protein, partial cds.
ACCESSION   AB014548
VERSION     AB014548.1  GI:3327109
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ishikawa,K., Nagase,T., Suyama,M., Miyajima,N., Tanaka,A.,
            Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes. X.
            The complete sequences of 100 new cDNA clones from brain which can
            code for large proteins in vitro
  JOURNAL   DNA Res. 5 (3), 169-176 (1998)
  MEDLINE   98403880
   PUBMED   9734811
REFERENCE   2  (bases 1 to 5177)
  AUTHORS   Ohara,O., Suyama,M., Nagase,T. and Ishikawa,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-1998) Osamu Ohara, Kazusa DNA Research Institute,
            Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
            292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913,
            Fax:+81-438-52-3914)
FEATURES             Location/Qualifiers
     source          1..5177
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HJ03994"
                     /sex="male"
                     /tissue_type="brain"
                     /clone_lib="pBluescriptII SK plus"
                     /dev_stage="adult"
     gene            1..5177
                     /gene="KIAA0648"
     CDS             <1..2558
                     /gene="KIAA0648"
                     /codon_start=3
                     /product="KIAA0648 protein"
                     /protein_id="BAA31623.1"
                     /db_xref="GI:3327110"
                     /translation="LYYLYASLDPNAVKALNEMWKCQNMLRSHVRELLDLHKQPTSEA
                     NCSAMFGKLMTIAKNLPDPGKAQDFVKKFNQVLGDDEKLRSQLELLISPTCSCKQADI
                     CVREIARKLANPKQPTNPFLEMVKFLLERIAPVHIDSEAISALVKLMNKSIEGTADDE
                     EEGVSPDTAIRSGLELLKVLSFTHPTSFHSAETYESLLQCLRMEDDKVAEAAIQIFRN
                     TGHKIETDLPQIRSTLIPILHQKAKRGTPHQAKQAVHCIHAIFTNKEVQLAQIFEPLS
                     RSLNADVPEQLITPLVSLGHISMLAPDQFASPMKSVVANFIVKDLLMNDRSTGEKNGK
                     LWSPDEEVSPEVLAKVQAIKLLVRWLLGMKNNQSKSANSTLRLLSAMLVSEGDLTEQK
                     RISKSDMSRLRLAAGSAIMKLAQEPCYHEIITPEQFQLCALVINDECYQVRQIFAQKL
                     HKALVKLLLPLEYMAIFALCAKDPVKERRAHARQCLLKNISIRREYIKQNPMATEKLL
                     SLLPEYVVPYMIHLLAHDPDFTRSQDVDQLRDIKECLWFMLEVLMTKNENNSHAFMKK
                     MAENIKLTRDAQSPDESKTNEKLYTVCDVALCVINSKSALCNADSPKDPVLPMKFFTQ
                     PEKDFCNDKSYISEETRVLLLTGKPKPAGVLGAVNKPLSATGRKPYVRSTGTETGSNI
                     NVNSELNPSTGNRSREQSSEAAETGVSENEENPVRIISVTPVKNIDPVKNKEINSDQA
                     TQGNISSDRGKKRTVTAAGAENIQQKTDEKVDESGPPAPSKPRRGRRPKSESQGNATK
                     NDDLNKPINKGRKRAAVGQESPGGLEAGNAKAPKLQDLAKKAAPAERQIDLQR"
BASE COUNT     1690 a    931 c   1037 g   1519 t
ORIGIN

  Query Match            25.0%;  Score 1041.4;  DB 9;  Length 5177;
  Best Local Similarity  68.0%;  Pred. No. 3.5e-204;
  Matches 1450;  Conservative    0;  Mismatches  681;  Indels    0;  Gaps    0;

Qy       1427 GCTTATATTACTTGTATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGT 1486
Db          1 GCTTATATTACTTATATGCTAGTTTGGATCCAAATGCTGTAAAAGCTCTCAACGAAATGT 60

Qy       1487 GGAAATGTCAAAATCTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAAC 1546
Db         61 GGAAGTGTCAGAACATGCTTCGGAGCCATGTACGCGAACTATTGGATTTGCACAAGCAGC 120

Qy       1547 CCAAAACAGATGCCAGTGTCAAGGCCATATTTCAAAAGTGATGGTTATTACAAGAAATT 1606
Db        121 CTACATCAGAGGCTAACTGTTCTGCCATGTTTGGAAAACTGATGACCATAGCAAAGAATT 180

Qy       1607 TACCTGATCCTGGTAAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATG 1666
Db        181 TGCCTGACCCCGGGAAAGCACAAGATTTTGTGAAGAAATTTAACCAGGTTCTCGGCGATG 240

Qy       1667 ATGAGAAAATAAGAAAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGG 1726
Db        241 ATGAGAAACTTCGGTCTCAGTTGGAGTTATTAATTAGCCCAACCTGTTCTTGCAAACAAG 300

Qy       1727 CTGAAGGTTGTGTGCGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATC 1786
Db        301 CAGATATTTGTGTGAGAGAAATAGCCCGGAAACTTGCAAATCCTAAGCAACCAACAAATC 360

Qy       1787 CTTTCCTGGAAATGATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCG 1846
Db        361 CTTTTCTAGAGATGGTCAAATTTCTGTTGGAAAGAATCGCACCTGTGCACATTGATTCAG 420

Qy       1847 AATCTATCAGTGCTCTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATG 1906
Db        421 AAGCCATAAGTGCACTAGTGAAATTGATGAATAAGTCAATAGAGGGGACAGCAGATGATG 480

Qy       1907 AAGATGAGGGTGTTCCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTAC 1966
Db        481 AAGAGGAGGGTGTAAGTCCAGATACAGCTATCCGTTCAGGACTTGAACTTCTTAAGGTTC 540

Qy       1967 TCTCATTTACACATCCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTT 2026
Db        541 TGTCTTTTACACATCCTACCTCGTTCCACTCTGCAGAGACATATGAGTCCTTGTTACAGT 600

Qy       2027 GTCTGAAAATGGATGATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAG 2086
Db        601 GCCTAAGAATGGAGGATGACAAGGTAGCAGAAGCTGCTATTCAAATTTTTAGAAATACAG 660

Qy       2087 GAAGCAAAATTGAAGAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATC 2146
Db        661 GTCACAAAATAGAAACAGACCTTCCCCAGATACGATCGACCTTAATTCCCATTTTACATC 720

Qy       2147 ACAAATCTAAAAAAGGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGA 2206
Db        721 AAAAAGCAAAGAGGGGTACTCCACACCAAGCAAAACAGGCTGTGCACTGTATACACGCCA 780

Qy       2207 TATTTTCTAGTAAAGAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAG 2266
Db        781 TATTCACAAATAAAGAAGTCCAGCTTGCACAGATTTTTGAGCCACTCAGTAGGAGTCTGA 840

Qy       2267 ATCCAAGCAACCTGGAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCC 2326
Db        841 ATGCTGATGTGCCAGAACAACTTATAACTCCATTAGTTTCATTGGGCCACATTTCTATGT 900

Qy       2327 TTGCACCTGATCAATTTGCTGCTCCTTGGAAATCTTGGGTAGCTACTTTCATTGTGAAAG 2386
Db        901 TAGCACCAGATCAGTTTGCTTCCCCAATGAAATCTGTAGTAGCAAATTTTATTGTGAAAG 960

Qy       2387 ATCTTCTCATGAATGATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATG 2446
Db        961 ATCTGCTAATGAATGACAGGTCAACAGGTGAAAAGAATGGAAAACTGTGGTCTCCAGATG 1020

Qy       2447 AAGAAGTATCTCCTGAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGC 2506
Db       1021 AAGAGGTTTCCCCTGAAGTACTAGCAAAGGTACAGGCAATTAAACTTCTGGTAAGGTGGC 1080

Qy       2507 TACTTGGAATGAAAAATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAA 2566
Db       1081 TGTTGGGTATGAAAAACAACCAGTCTAAATCTGCCAATTCAACCCTTCGGTTATTATCAG 1140

Qy       2567 CAATATTGCATAGTGATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGT 2626
```

```
Db        1141 CGATGTTGGTTAGTGAGGGTGACCTGACAGAGCAAAAGAGGATCAGTAAATCTGATATGT 1200

Qy        2627 CACGTCTGAGACTTGCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACC 2686
Db        1201 CTCGCTTGCGATTAGCTGCTGGTAGTGCCATAATGAAGCTTGCTCAGGAACCTTGTTACC 1260

Qy        2687 ATGAAATCATCACATTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCT 2746
Db        1261 ATGAAATTATTACCCCAGAACAGTTTCAGCTCTGTGCACTTGTTATTAATGATGAGTGTT 1320

Qy        2747 ATCAAGTAAGACAAGTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTC 2806
Db        1321 ACCAAGTAAGGCAGATATTTGCTCAGAAGCTGCATAAGGCACTTGTGAAGTTACTGCTCC 1380

Qy        2807 CACTTGAGTATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAG 2866
Db        1381 CATTGGAGTATATGGCGATCTTTGCCTTGTGTGCCAAAGATCCTGTGAAGGAGAGAAGAG 1440

Qy        2867 CTCATGCTAGGCAATGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGC 2926
Db        1441 CACACGCACGACAATGTTTACTGAAAAATATCAGTATACGCAGGGAATACATTAAGCAGA 1500

Qy        2927 ATGCAGCTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAA 2986
Db        1501 ATCCTATGGCTACTGAGAAATTATTATCACTGTTGCCTGAATATGTAGTTCCATACATGA 1560

Qy        2987 TTCACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAG 3046
Db        1561 TTCACCTGCTAGCCCATGATCCAGATTTTACAAGATCACAAGATGTTGATCAGCTTCGTG 1620

Qy        3047 ATGTTAAAGAATGTCTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACA 3106
Db        1621 ATATCAAAGAGTGCCTATGGTTCATGCTTGAAGTTTTAATGACAAAGAATGAAAACAATA 1680

Qy        3107 GTCACGCTTTTATCAGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGAC 3166
Db        1681 GCCATGCCTTTATGAAGAAGATGGCAGAGAACATCAAGTTAACCAGAGATGCCCAGTCTC 1740

Qy        3167 CAGATGATGCAAAAATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCA 3226
Db        1741 CAGATGAATCCAAGACAAATGAAAAACTGTATACAGTATGTGATGTGGCTCTCTGTGTTA 1800

Qy        3227 TCATGTCAAAGAGTACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTC 3286
Db        1801 TAAATAGTAAAAGTGCTTTGTGCAATGCAGATTCACCAAAGGACCCAGTCCTCCCAATGA 1860

Qy        3287 GTTTCTTCACTCAACCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAA 3346
Db        1861 AATTTTTACACAACCTGAAAAGGACTTCTGTAACGATAAGAGTTATATTTCAGAAGAGA 1920

Qy        3347 TGAAATCATTTTTCACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACA 3406
Db        1921 CAAGAGTACTTCTGTTAACAGGAAAGCCAAAGCCTGCTGGAGTACTAGGTGCAGTAAATA 1980

Qy        3407 AGCCACTTTCATCAGCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAA 3466
Db        1981 AGCCTTTATCAGCAACGGGAAGGAAACCCTATGTTAGAAGCACTGGCACTGAGACTGGAA 2040

Qy        3467 GCAATGCAAGCAGCAGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATA 3526
Db        2041 GCAATATTAATGTAAATTCAGAGCTGAACCCTTCAACCGGAAATCGATCAAGGGAACAGA 2100

Qy        3527 GTTCTGAAATGGATCACAGTGAAAATGAAGA 3557
Db        2101 GTTCAGAGGCAGCAGAAACTGGAGTTAGTGA 2131
```

RESULT 10
BC021408

```
LOCUS       BC021408                1976 bp    mRNA    linear   ROD 16-APR-2003
DEFINITION  Mus musculus RIKEN cDNA 9030416H16 gene, mRNA (cDNA clone MGC:29439
            IMAGE:3964501), complete cds.
ACCESSION   BC021408
VERSION     BC021408.1  GI:18204099
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1976)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1976)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 39 Row: b Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: GenomeScan gene
            prediction.
FEATURES             Location/Qualifiers
     source          1..1976
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="MGC:29439 IMAGE:3964501"
                     /tissue_type="Mammary tumor. Metallothionien-TGF alpha
                     model. 10 month old virgin mouse. Taken by biopsy."
                     /clone_lib="NCI_CGAP_Mam1"
```

```
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1..1976
                     /gene="9030416H16Rik"
                     /db_xref="LocusID:71521"
                     /db_xref="MGI:1918771"
     CDS             130..1884
                     /codon_start=1
                     /product="RIKEN cDNA 9030416H16"
                     /protein_id="AAH21408.1"
                     /db_xref="GI:18204100"
                     /db_xref="LocusID:71521"
                     /translation="MFGKLMTIAKNLPDPGKAQDFVKKFNQVLGDDEKLRSQLELLIS
                     PTCSCKQADVCVREIARKLANPKQPTNPFLEMVKFLLERIAPVHIDSEAISALVKLMN
                     KSIEGTADDEEEGVSPDSAIRSGLELLKVLSFTHPTSFHSAETYESLLQCLRMEDDKV
                     AEAAIQIFRNTGHKIETDLPQIRSTLIPILHQKAKRGTPHQAKQAVHCIHAIFSNKEV
                     QLAQIFEPLSRSLNADVPEQLITPLVSLGHISMLAPDQFASPMKSVVANFIVKDLLMN
                     DRSTGEKNGKLWSPDEEVSPEVLAKVQAIKLLVRWLLGMKNNQSKSANSTLRLLSAML
                     VSEGDLTEQKRISKSDMSRLRLAAGSAIMKLAQEPCYHEIITPEQFQLCALVINDECY
                     QVRQIFAQKLHKALVKLLLPLEYMAIFALCAKDPVKERRAHARQCLLKNISIRREYIK
                     QNPMATEKLLSLLPEYVVPYMIHLLAHDPDFTRSQDVDQLRDIKECLWFMLEVLMTKN
                     ENNSHAFMKKMAENIKLTRDAQSPDESKTNEKLYTVCDVALCVINSKSALCNADSPKD
                     PVLPMKFFTQPEKVISLT"
BASE COUNT      611 a    429 c    432 g    504 t
ORIGIN

  Query Match             22.5%;  Score 937.6;  DB 10;  Length 1976;
  Best Local Similarity   68.4%;  Pred. No. 1e-182;
  Matches 1297;  Conservative    0;  Mismatches  599;  Indels    0;  Gaps    0;

Qy       1468 AAAGCATTGAATGAAATGTGGAAATGTCAAAATCTGCTCCGACATCAAGTAAAGGATTTG 1527
Db         25 AAAGCTCTCAATGAAATGTGGAAATGTCAGAACATGCTTCGAAGTCATGTGCGTGAACTA 84

Qy       1528 CTTGACTTGATTAAGCAACCCAAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGTG 1587
Db         85 TTGGACTTACACAAGCAGCCTACATCAGAAGCGAACTGTTCTGCCATGTTTGGGAAACTG 144

Qy       1588 ATGGTTATTACAAGAAATTTACCTGATCCTGGTAAGGCTCAGGATTTCATGAAGAAATTC 1647
Db        145 ATGACCATAGCAAAGAATTTGCCTGACCCTGGAAAAGCACAAGATTTTGTAAAGAAATTT 204

Qy       1648 ACACAGGTGTTAGAAGATGATGAGAAAATAAGAAAGCAGTTAGAAGTACTTGTTAGTCCA 1707
Db        205 AACCAGGTTCTTGGTGATGATGAGAAACTGAGGTCTCAGTTGGAATTATTAATCAGCCCA 264

Qy       1708 ACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTGCGTGAAATAACTAAGAAGTTGGGCAAC 1767
Db        265 ACCTGTTCATGCAAGCAGGCTGACGTTTGTGTGAGGGAAATAGCTCGGAAACTTGCAAAT 324

Qy       1768 CCCAAACAGCCTACAAATCCTTTCCTGGAAATGATCAAGTTTCTCTTGGAGAGGATAGCA 1827
Db        325 CCTAAGCAGCCAACCAACCCTTTTCTAGAGATGGTCAAATTTCTTTTGGAAAGAATTGCT 384

Qy       1828 CCTGTGCACATAGATACCGAATCTATCAGTGCTCTTATTAAACAAGTGAACAAATCAATA 1887
Db        385 CCTGTGCACATCGATTCAGAAGCCATAAGTGCACTGGTAAAACTGATGAATAAGTCAATC 444

Qy       1888 GATGGAACAGCAGATGATGAAGATGAGGGTGTTCCAACTGATCAAGCCATCAGAGCAGGT 1947
Db        445 GAAGGGACGGCCGATGATGAAGAGGAGGGTGTCAGTCCAGACTCAGCCATCCGCTCAGGA 504

Qy       1948 CTTGAACTGCTTAAGGTACTCTCATTTACACATCCCATCTCATTTCATTCTGCTGAAACA 2007
Db        505 CTTGAGCTTCTTAAGGTTCTGTCTTTCACACATCCTACCTCGTTCCACTCTGCAGAGACA 564

Qy       2008 TTTGAATCATTACTGGCTTGTCTGAAAATGGATGATGAAAAAGTAGCAGAAGCTGCACTA 2067
Db        565 TATGAGTCCTTGTTACAGTGCCTAAGAATGGAGGATGACAAGGTAGCAGAAGCTGCAATA 624

Qy       2068 CAAATTTTCAAAAACACAGGAAGCAAAATTGAAGAGGATTTTCCACACATCAGATCAGCC 2127
Db        625 CAAATTTTAGAAACACAGGCCACAAAATAGAAACTGACCTTCCCCAGATACGGTCAACC 684

Qy       2128 TTGCTTCCTGTTTTACATCACAAATCTAAAAAAGGACCCCCCGTCAAGCCAAATATGCC 2187
Db        685 TTGATCCCCATTTTACATCAGAAAGCCAAGCGGGGCACTCCACACCAAGCAAAGCAGGCT 744

Qy       2188 ATTCATTGTATCCATGCGATATTTTCTAGTAAAGAGACCCAGTTTGCACAGATATTTGAG 2247
Db        745 GTTCACTGCATCCATGCCATCTTCTCAAACAAGGAGGTCCAGCTGGCACAGATTTTTGAG 804

Qy       2248 CCTCTGCATAAGAGCCTAGATCCAAGCAACCTGGAACATCTCATAACACCATTGGTTACT 2307
Db        805 CCACTCAGTAGGAGTCTGAATGCTGATGTACCAGAGCAACTTATAACTCCATTAGTTTCA 864

Qy       2308 ATTGGTCATATTGCTCTCCTTGCACCTGATCAATTTGCTGCTCCTTGGAAATCTTGGGTA 2367
Db        865 CTGGGCCACATTTCCATGTTAGCACCAGATCAGTTTGCCTCCCCGATGAAATCTGTAGTG 924

Qy       2368 GCTACTTTCATTGTGAAAGATCTTCTCATGAATGATCGGCTTCCAGGGAAAAAGACAACT 2427
Db        925 GCAAACTTTATTGTTAAAGACCTTCTAATGAACGACAGGTCAACAGGTGAGAAGAATGGA 984

Qy       2428 AAACTTTGGGTTCCAGATGAAGAAGTATCTCCTGAGACAATGGTCAAAATTCAGGCTATT 2487
Db        985 AAATTATGGTCTCCAGATGAGGAAGTTTCTCCTGAAGTTCTAGCAAAGGTACAGGCAATT 1044

Qy       2488 AAAATGATGGTTCGATGGCTACTTGGAATGAAAAATAATCACAGTAAATCAGGAACTTCT 2547
Db       1045 AAACTTCTGGTAAGGTGGCTGTTGGGTATGAAAAACAACCAGTCTAAATCTGCCAACTCA 1104

Qy       2548 ACCTTAAGATTGCTAACAACAATATTGCATAGTGATGGAGACTTGACAGAACAGGGGAAA 2607
Db       1105 ACTCTTCGGTTATTATCAGCCATGCTGGTTAGTGAGGGCGACCTGACAGAGCAAAAGAGG 1164

Qy       2608 ATTAGTAAACCAGATATGTCACGTCTGAGACTTGCTGCTGGGAGTGCTATTGTGAAGCTG 2667
Db       1165 ATCAGTAAATCTGATATGTCTCGCTTGCGATTAGCTGCTGGTAGTGCCATAATGAAGCTT 1224

Qy       2668 GCACAAGAACCCTGTTACCATGAAATCATCACATTAGAACAATATCAGCTATGTGCATTA 2727
Db       1225 GCTCAGGAACCTTGCTACCATGAAATTATTACCCCAGAACAGTTTCAGCTCTGTGCACTG 1284

Qy       2728 GCTATCAACGATGAATGCTATCAAGTAAGACAAGTGTTTGCCCAGAAACTTCACAAAGGC 2787
Db       1285 GTTATTAATGATGAGTGCTACCAAGTAAGGCAGATATTTGCCCAGAAGCTTCATAAGGCA 1344

Qy       2788 CTTTCCCGTTTACGGCTTCCACTTGAGTATATGGCAATCTGTGCCCTTTGTGCAAAAGAT 2847
Db       1345 CTTGTGAAGTTGCTTCTCCCACTGGAGTATATGGCCATCTTTGCTTTGTGTGCCAAAGAC 1404

Qy       2848 CCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAATGTTTGGTGAAAAATATAAATGTAAGG 2907
Db       1405 CCTGTGAAGGAAAGGAGAGCACACGCTCGGCAGTGTTTGTTAAAGAACATCAGCATCCGC 1464

Qy       2908 CGGGAGTATCTGAAGCAGCATGCAGCTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAG 2967
Db       1465 AGGGAGTACATCAAACAGAACCCCATGGCCACTGAGAAATTATTGTCACTGCTGCCTGAA 1524

Qy       2968 TATGTTGTTCCATATACAATTCACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAG 3027
Db       1525 TATGTGGTTCCATACATGATTCACCTCCTAGCCCATGATCCTGATTTCACACGGTCACAA 1584

Qy       3028 GATATTGAACAACTTAAAGATGTTAAAGAATGTCTTTGGTTTGTTCTGGAAATATTAATG 3087
Db       1585 GATGTTGATCAACTTCGTGATATAAAAGAGTGCCTGTGGTTTATGCTTGAAGTCTTAATG 1644

Qy       3088 GCTAAAAATGAAAATAACAGTCACGCTTTTATCAGAAAGATGGTAGAAAATATTAAACAA 3147
Db       1645 ACAAAGAATGAAAACAACAGCCATGCATTCATGAAAAAGATGGCAGAGAATATCAAGCTA 1704

Qy       3148 ACAAAAGATGCCCAAGGACCAGATGATGCAAAAATGAATGAAAAACTGTACACTGTGTGT 3207
Db       1705 ACCAGAGACGCCCAGTCTCCCGATGAATCCAAGACAAATGAAAAACTTTATACGGTTTGT 1764
```

```
Qy        3208 GATGTTGCCATGAATATCATCATGTCAAAGAGTACTACATACAGTTTGGAATCTCCTAAA 3267

Db        1765 GACGTGGCTCTGTGTGTTATAAATAGTAAAAGTGCTTTGTGCAATGCAGACTCACCAAAG 1824

Qy        3268 GACCCGGTACTACCAGCTCGTTTCTTCACTCAACCTGACAAGAATTTCAGTAACACCAAA 3327

Db        1825 GACCCAGTCCTCCCAATGAAGTTCTTTACACAGCCTGAAAAGGTAATTTCCCTCACATAA 1884

Qy        3328 AATTATCTGCCTCCTGAAATGAAATCATTTTTCACT 3363

Db        1885 TTTCTCTGTCATAGTAAAAGTGATTTAAATTTCAGT 1920


RESULT 11
AK098331
LOCUS       AK098331                2265 bp    mRNA    linear   PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ41012 fis, clone UTERU2018523, highly similar
            to Human androgen-induced prostate proliferative shutoff associated
            protein (AS3) mRNA.
ACCESSION   AK098331
VERSION     AK098331.1  GI:21758323
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kawakami,B., Sugiyama,A., Takemoto,M., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.
            and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2265)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1..2265
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UTERU2018523"
                     /tissue_type="uterus"
                     /clone_lib="UTERU2"
                     /note="cloning vector: pME18SFL3"
     CDS             125..1927
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAC05286.1"
                     /db_xref="GI:21758324"
                     /translation="MDFTAQPKPATALCGVVSADGKIAYPPGVKEITDKITTDEMIKR
                     LKMVVKTFMDMDQDSEDEKQQYLPLALHLASEFFLRNPNKDVRLLVACCLADIFRIYA
                     PEAPYTSHDKLKDIFLFITRQLKGLEDTKSPQFNRYFYLLENLAWVKSYNICFELEDC
                     NEIFIQLFRTLFSVINNSHNKKVQMHMLDLMSSIIMEGDGVTQELLDSILINLIPAHK
                     NLNKQSFDLAKVLLKRTVQTIEACIANFFNQVLVLGRSSVSDLSEHVFDLIQELFAID
                     PHLLLSVMPQLEFKLKSNDGEERLAVVRLLAKLFGSKDSDLATQNRPLWQCFLGRFND
                     IHVPVRLESVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVTIITAAKRDL
                     ALVNDQLLGFVRERTLDKRWRVRKEAMMGLAQLYKKYCLHGEAGKEAAEKVSWIKDKL
                     LHIYYQNSIDDKLLVEKIFAQYLVPHNLETEERMKCLYYLYASLDPNAVKALNEMWKC
                     QNMLRSHVRELLDLHKQPTSEANCSAMFGKLMTIAKNLPDPGKAQDFVKKFNQVLGDD
                     EKLRSQLELLISPTCSCKQADICVVSKSYFTLFL"
BASE COUNT      696 a    427 c    476 g    666 t
ORIGIN

  Query Match          22.3%;  Score 932.2;  DB 9;  Length 2265;
  Best Local Similarity   71.5%;  Pred. No. 1.3e-181;
  Matches 1225;  Conservative    0;  Mismatches  488;  Indels    0;  Gaps    0;

Qy          28 GATGGAAAAATTACATATCCGCCTGGGGTCAAGGAAATATCAGATAAAATATCTAAAGAG 87

Db         182 GACGGGAAGATCGCTTACCCTCCGGGGGTAAAAGAGATCACCGACAAGATCACCACGGAC 241

Qy          88 GAGATGGTGAGACGATTAAAGATGGTTGTGAAAACTTTTATGGATATGGACCAGGACTCT 147

Db         242 GAGATGATCAAACGCCTGAAGATGGTAGTGAAAACCTTTATGGATATGGATCAGGACTCA 301

Qy         148 GAAGAAGAAAAGGAGCTTTATTTAAACCTAGCTTTACATCTTGCTTCAGATTTTTTCTC 207

Db         302 GAAGATGAAAAACAGCAGTATCTCCCACTAGCCTTGCATCTTGCATCTGAATTCTTCCTC 361

Qy         208 AAGCATCCTGGTAAAGATGTTCGCTTACTGGTAGCCTGCTGCCTTGCTGATATTTTCAGG 267

Db         362 AGGAACCCCAATAAAGATGTGCGTCTCCTTGTAGCATGTTGTTTGGCTGATATCTTTCGT 421

Qy         268 ATTTATGCTCCTGAAGCTCCTTACACATCCCCTGATAAACTAAAGGATATATTTATGTTT 327

Db         422 ATCTATGCCCCAGAAGCTCCATATACTTCCCATGATAAACTTAAGGACATATTTTTGTTT 481

Qy         328 ATAACAAGACAGTTGAAGGGGCTAGAGGATACAAAGAGCCCACAATTCAATAGGTATTTT 387

Db         482 ATTACCAGACAATTAAAAGGTTTGGAGGATACAAAGAGTCCACAGTTTAATAGATACTTT 541

Qy         388 TATTTACTTGAGAACATTGCTTGGGTCAAGTCATATAACATATGCTTTGAGTTAGAAGAT 447

Db         542 TATTTATTAGAGAATTTAGCTTGGGTTAAATCATATAACATCTGCTTTGAATTGGAAGAT 601

Qy         448 AGCAATGAAATTTTCACCCAGCTATACAGAACCTTATTTTCAGTTATAAACAATGGCCAC 507

Db         602 TGCAATGAAATTTTTATTCAGCTTTTTAGAACTCTCTTCTCAGTGATCAACAATAGCCAC 661

Qy         508 AATCAGAAAGTCCATATGCACATGGTAGACCTTATGAGCTCTATTATTTGTGAAGGTGAT 567

Db         662 AATAAGAAGGTACAAATGCACATGCTAGATTTGATGAGTTCTATCATCATGGAAGGTGAT 721

Qy         568 ACAGTGTCTCAGGAGCTTTTGGATACGGTTTTAGTAAATCTGGTACCTGCTCATAAGAAT 627

Db         722 GGAGTTACTCAAGAATTATTGGACTCCATTCTTATTAACCTCATTCCTGCACATAAGAAC 781

Qy         628 TTAAACAAGCAAGCATATGATTTGGCAAAGGCTTTACTGAAGAGGACAGCTCAAGCTATT 687

Db         782 TTAAATAAACAGTCCTTTGACCTTGCAAAAGTGCTATTGAAAAGAACAGTCCAGACTATT 841

Qy         688 GAGCCATATATTACCACTTTTTTTAATCAGGTTCTGATGCTTGGGAAAACATCTATCAGC 747

Db         842 GAGGCATGCATTGCTAATTTTTTCAATCAAGTCCTGGTGCTGGGAAGATCATCAGTAAGT 901

Qy         748 GATTTGTCAGAGCATGTCTTTGACTTAATTTTGGAGCTCTACAATATTGATAGTCATTTG 807

Db         902 GATTTGTCAGAACATGTATTTGATCTGATTCAGGAACTTTTTGCTATAGATCCTCATTTA 961

Qy         808 CTGCTCTCTGTTTTACCCCAGCTTGAATTTAAATTAAAGAGCAATGATAATGAGGAGCGC 867

Db         962 TTATTATCCGTCATGCCACAGCTTGAATTCAAACTAAAGAGCAATGATGGAGAAGAGCGA 1021

Qy         868 CTACAAGTTGTTAAACTACTGGCAAAAATGTTTGGGGCAAAGGATTCAGAATTGGCTTCT 927

Db        1022 TTAGCTGTTGTTCGACTTCTAGCTAAATTGTTTGGCTCCAAAGATTCTGATTTGGCAACA 1081

Qy         928 CAAAACAAGCCACTTTGGCAGTGCTACTTGGGCAGGTTTAATGATATCCATGTACCAATC 987

Db        1082 CAGAATCGTCCTCTTTGGCAATGTTTTCTTGGACGATTTAATGATATTCATGTTCCTGTG 1141
```

```
Qy        988 CGCCTGGAATGTGTGAAATTTGCTAGCCATTGTCTCATGAACCATCCTGATTTAGCAAAA 1047
               |  | |||  |||||||||||| ||  |||||| |  ||||| || || |||||||| ||
Db       1142 AGATTAGAAAGTGTGAAATTTGCCAGTCATTGTTTAATGAATCACCCAGATTTAGCGAAG 1201

Qy       1048 GACTTAACAGAGTATCTTAAAGTGAGGTCACATGACCCTGAGGAAGCTATTAGACATGAT 1107
              ||  |  ||||| |||  |  || || || ||||||||  || || |||||||||  | ||||||
Db       1202 GATCTCACAGAATATTTAAAGGTTAGATCACATGATCCAGAAGAAGCTATTCGTCATGAT 1261

Qy       1108 GTTATTGTGTCAATAGTTACAGCTGCTAAAAAGGATATTCTTCTGGTCAATGATCACTTA 1167
              || |||||   |  |||  |  ||||||||  || |  |||   |       |  ||  ||||||||   |
Db       1262 GTCATTGTTACTATAATAACAGCTGCCAAGAGGGACCTGGCCTTAGTAAATGATCAGCTG 1321

Qy       1168 CTTAATTTTGTGAGAGAGAGAACATTAGACAAACGATGGAGAGTACGCAAAGAAGCCATG 1227
              |||    ||||| || || |||||| |  || ||||| ||| ||||| | |||||||| |||
Db       1322 CTTGGCTTTGTAAGGGAAAGAACACTGGATAAACGGTGGCGAGTAAGAAAAGAAGCTATG 1381

Qy       1228 ATGGGACTTGCCCAAATTTATAAGAAATATGCTTTACAGTCAGCAGCTGGAAAAGATGCT 1287
              ||||| || || ||  ||||||||||||      |  |  ||        | ||| ||||| || |||
Db       1382 ATGGGTCTGGCTCAGCTTTATAAGAAATACTGTCTTCATGGTGAAGCAGGAAAGGAAGCT 1441

Qy       1288 GCAAAACAGATAGCATGGATCAAAGACAAATTGCTACATATATATTATCAAAATAGTATT 1347
              ||| |  |   |   ||||| || |||||  | || ||||||  |||||||| || |||
Db       1442 GCAGAGAAAGTCAGCTGGATAAAGGACAAACTTCTGCATATTTATTATCAGAACAGCATT 1501

Qy       1348 GATGATCGACTACTTGTTGAACGGATCTTTGCTCAATACATGGTTCCTCACAATTTAGAA 1407
              || ||   |||  | || ||     |||||||||| ||   | || || |||||   |  |||
Db       1502 GACGACAAACTGTTGGTAGAGAAAATCTTTGCTCAGTATCTTGTCCCCCACAACCTGGAA 1561

Qy       1408 ACTACAGAACGGATGAAATGCTTATATTACTTGTATGCCACACTGGATTTAAATGCTGTG 1467
              ||    |||  |  ||||||||||||||||||| |||||  |     |||||  ||||||||
Db       1562 ACAGAAGAGAGAATGAAATGCTTATATTACTTATATGCTAGTTTGGATCCAAATGCTGTA 1621

Qy       1468 AAAGCATTGAATGAAATGTGGAAATGTCAAAATCTGCTCCGACATCAAGTAAAGGATTTG 1527
              |||||  |  ||  ||||||||||| ||||| ||    ||  |||     || ||       ||   |
Db       1622 AAAGCTCTCAACGAAATGTGGAAGTGTCAGAACATGCTTCGGAGCCATGTACGCGAACTA 1681

Qy       1528 CTTGACTTGATTAAGCAACCCAAAACAGATGCCAGTGTCAAGGCCATATTTCAAAAGTG 1587
               | || |||    ||||| || |  | |||| || |           ||||| |||   |||| ||
Db       1682 TTGGATTTGCACAAGCAGCCTACATCAGAGGCTAACTGTTCTGCCATGTTTGGAAAACTG 1741

Qy       1588 ATGGTTATTACAAGAAATTTACCTGATCCTGGTAAGGCTCAGGATTTCATGAAGAAATTC 1647
              |||    ||   |||  ||||| ||||| || || || || || ||  |||||   |||||||||
Db       1742 ATGACCATAGCAAAGAATTTGCCTGACCCCGGGAAAGCACAAGATTTTGTGAAGAAATTT 1801

Qy       1648 ACACAGGTGTTAGAAGATGATGAGAAAATAAGAAAGCAGTTAGAAGTACTTGTTAGTCCA 1707
              |  |||||   | |   |||||||||||| |    |     ||||| ||   || |   |||| |||
Db       1802 AACCAGGTTCTCGGCGATGATGAGAAACTTCGGTCTCAGTTGGAGTTATTAATTAGCCCA 1861

Qy       1708 ACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1740
              || || || ||||| || || ||    |||||||
Db       1862 ACCTGTTCTTGCAAACAAGCAGATATTTGTGTG 1894
```

RESULT 12
BC041361

```
LOCUS       BC041361                2705 bp    mRNA    linear   PRI 09-JAN-2003
DEFINITION  Homo sapiens, Similar to KIAA0648 protein, clone MGC:43818
            IMAGE:5273075, mRNA, complete cds.
ACCESSION   BC041361
VERSION     BC041361.1  GI:27552793
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2705)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 75 Row: c Column: 19.
FEATURES             Location/Qualifiers
     source          1..2705
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:43818 IMAGE:5273075"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
     CDS             545..2347
                     /codon_start=1
                     /product="Similar to KIAA0648 protein"
                     /protein_id="AAH41361.1"
                     /db_xref="GI:27552794"
                     /translation="MDFTAQPKPATALCGVVSADGKIAYPPGVKEITDKITTDEMIKR
                     LKMVVKTFMDMDQDSEDEKQQYLPLALHLASEFFLRNPNKDVRLLVACCLADIFRIYA
                     PEAPYTSHDKLKDIFLFITRQLKGLEDTKSPQFNRYFYLLENLAWVKSYNICFELEDC
                     NEIFIQLFRTLFSVINNSHNKKVQMHMLDLMSSIIMEGDGVTQELLDSILINLIPAHK
                     NLNKQSFDLAKVLLKRTVQTIEACIANFFNQVLVLGRSSVSDLSEHVFDLIQELFAID
                     PHLLLSVMPQLEFKLKSNDGEERLAVVRLLAKLFGSKDSDLATQNRPLWQCFLGRFND
                     IHVPVRLESVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVTIITAAKRDL
                     ALVNDQLLGFVRERTLDKRWRVRKEAMMGLAQLYKKYCLHGEAGKEAAEKVSWIKDKL
                     LHIYYQNSIDDKLLVEKIFAQYLVPHNLETEERMKCLYYLYASLDPNAVKALNEMWKC
                     QNMLRSHVRELLDLHKQPTSEANCSAMFGKLMTIAKNLPDPGKAQDFVKKFNQVLGDD
                     EKLRSQLELLISPTCSCKQADICVVSKSYFTLFL"
BASE COUNT      751 a    589 c    654 g    711 t
ORIGIN

  Query Match             22.3%;  Score 932.2;  DB 9;  Length 2705;
  Best Local Similarity   71.5%;  Pred. No. 1.3e-181;
  Matches 1225;  Conservative    0;  Mismatches  488;  Indels    0;  Gaps    0;

Qy         28 GATGGAAAAATTACATATCCGCCTGGGGTCAAGGAAATATCAGATAAAATATCTAAAGAG 87
              || || || ||  | || || || ||||| || || ||  | || || ||  | | `||
Db        602 GACGGGAAGATCGCTTACCCTCCGGGGGTAAAAGAGATCACCGACAAGATCACCACGGAC 661

Qy         88 GAGATGGTGAGACGATTAAAGATGGTTGTGAAAACTTTTATGGATATGGACCAGGACTCT 147
              |||||| |  | |||  | |||||||| ||||||| ||||||||||||||| |||||||
Db        662 GAGATGATCAAACGCCTGAAGATGGTAGTGAAAACCTTTATGGATATGGATCAGGACTCA 721

Qy        148 GAAGAAGAAAAGGAGCTTTATTTAAACCTAGCTTTACATCTTGCTTCAGATTTTTTCTC 207
              ||||| |||||  |||  |||  |       ||||| || |||||||| || || || || |||
Db        722 GAAGATGAAAAACAGCAGTATCTCCCACTAGCCTTGCATCTTGCATCTGAATTCTTCCTC 781

Qy        208 AAGCATCCTGGTAAAGATGTTCGCTTACTGGTAGCCTGCTGCCTTGCTGATATTTTCAGG 267
              | | | ||    ||||||||| ||   | || ||||| || ||   | |||||||| ||   |
Db        782 AGGAACCCCAATAAAGATGTGCGTCTCCTTGTAGCATGTTGTTTGGCTGATATCTTTCGT 841

Qy        268 ATTTATGCTCCTGAAGCTCCTTACACATCCCCTGATAAACTAAAGGATATATTTATGTTT 327
              || ||||| || |||||||| || || |||| ||||||||| ||||| |||||| |||||
Db        842 ATCTATGCCCCAGAAGCTCCATATACTTCCCATGATAAACTTAAGGACATATTTTGTTT 901
```

```
Qy       328 ATAACAAGACAGTTGAAGGGGCTAGAGGATACAAAGAGCCCACAATTCAATAGGTATTTT 387
Db       902 ATTACCAGACAATTAAAAGGTTTGGAGGATACAAAGAGTCCACAGTTTAATAGATACTTT 961

Qy       388 TATTTACTTGAGAACATTGCTTGGGTCAAGTCATATAACATATGCTTTGAGTTAGAAGAT 447
Db       962 TATTTATTAGAGAATTTAGCTTGGGTTAAATCATATAACATCTGCTTTGAATTGGAAGAT 1021

Qy       448 AGCAATGAAATTTTCACCCAGCTATACAGAACCTTATTTTCAGTTATAAACAATGGCCAC 507
Db      1022 TGCAATGAAATTTTTATTCAGCTTTTTAGAACTCTCTTCTCAGTGATCAACAATAGCCAC 1081

Qy       508 AATCAGAAAGTCCATATGCACATGGTAGACCTTATGAGCTCTATTATTTGTGAAGGTGAT 567
Db      1082 AATAAGAAGGTACAAATGCACATGCTAGATTTGATGAGTTCTATCATCATGGAAGGTGAT 1141

Qy       568 ACAGTGTCTCAGGAGCTTTTGGATACGGTTTTAGTAAATCTGGTACCTGCTCATAAGAAT 627
Db      1142 GGAGTTACTCAAGAATTATTGGACTCCATTCTTATTAACCTCATTCCTGCACATAAGAAC 1201

Qy       628 TTAAACAAGCAAGCATATGATTTGGCAAAGGCTTTACTGAAGAGGACAGCTCAAGCTATT 687
Db      1202 TTAAATAAACAGTCCTTTGACCTTGCAAAAGTGCTATTGAAAAGAACAGTCCAGACTATT 1261

Qy       688 GAGCCATATATTACCACTTTTTTTAATCAGGTTCTGATGCTTGGGAAAACATCTATCAGC 747
Db      1262 GAGGCATGCATTGCTAATTTTTTCAATCAAGTCCTGGTGCTGGGAAGATCATCAGTAAGT 1321

Qy       748 GATTTGTCAGAGCATGTCTTTGACTTAATTTTGGAGCTCTACAATATTGATAGTCATTTG 807
Db      1322 GATTTGTCAGAACATGTATTTGATCTGATTCAGGAACTTTTTGCTATAGATCCTCATTTA 1381

Qy       808 CTGCTCTCTGTTTTACCCCAGCTTGAATTTAAATTAAAGAGCAATGATAATGAGGAGCGC 867
Db      1382 TTATTATCCGTCATGCCACAGCTTGAATTCAAACTAAAGAGCAATGATGGAGAAGAGCGA 1441

Qy       868 CTACAAGTTGTTAAACTACTGGCAAAAATGTTTGGGGCAAAGGATTCAGAATTGGCTTCT 927
Db      1442 TTAGCTGTTGTTCGACTTCTAGCTAAATTGTTTGGCTCCAAAGATTCTGATTTGGCAACA 1501

Qy       928 CAAAACAAGCCACTTTGGCAGTGCTACTTGGGCAGGTTTAATGATATCCATGTACCAATC 987
Db      1502 CAGAATCGTCCTCTTTGGCAATGTTTTCTTGGACGATTTAATGATATTCATGTTCCTGTG 1561

Qy       988 CGCCTGGAATGTGTGAAATTTGCTAGCCATTGTCTCATGAACCATCCTGATTTAGCAAAA 1047
Db      1562 AGATTAGAAAGTGTGAAATTTGCCAGTCATTGTTTAATGAATCACCCAGATTTAGCGAAG 1621

Qy      1048 GACTTAACAGAGTATCTTAAAGTGAGGTCACATGACCCTGAGGAAGCTATTAGACATGAT 1107
Db      1622 GATCTCACAGAATATTTAAAGGTTAGATCACATGATCCAGAAGAAGCTATTCGTCATGAT 1681

Qy      1108 GTTATTGTGTCAATAGTTACAGCTGCTAAAAAGGATATTCTTCTGGTCAATGATCACTTA 1167
Db      1682 GTCATTGTTACTATAATAACAGCTGCCAAGAGGGACCTGGCCTTAGTAAATGATCAGCTG 1741

Qy      1168 CTTAATTTTGTGAGAGAGAGAACATTAGACAAACGATGGAGAGTACGCAAAGAAGCCATG 1227
Db      1742 CTTGGCTTTGTAAGGGAAAGAACACTGGATAAACGGTGGCGAGTAAGAAAAGAAGCTATG 1801

Qy      1228 ATGGGACTTGCCCAAATTTATAAGAAATATGCTTTACAGTCAGCAGCTGGAAAAGATGCT 1287
Db      1802 ATGGGTCTGGCTCAGCTTTATAAGAAATACTGTCTTCATGGTGAAGCAGGAAAGGAAGCT 1861

Qy      1288 GCAAAACAGATAGCATGGATCAAAGACAAATTGCTACATATATATTATCAAAATAGTATT 1347
Db      1862 GCAGAGAAAGTCAGCTGGATAAAGGACAAACTTCTGCATATTTATTATCAGAACAGCATT 1921

Qy      1348 GATGATCGACTACTTGTTGAACGGATCTTTGCTCAATACATGGTTCCTCACAATTTAGAA 1407
Db      1922 GACGACAAACTGTTGGTAGAGAAAATCTTTGCTCAGTATCTTGTCCCCCACAACCTGGAA 1981

Qy      1408 ACTACAGAACGGATGAAATGCTTATATTACTTGTATGCCACACTGGATTTAAATGCTGTG 1467
Db      1982 ACAGAAGAGAGAATGAAATGCTTATATTACTTATATGCTAGTTTGGATCCAAATGCTGTA 2041

Qy      1468 AAAGCATTGAATGAAATGTGGAAATGTCAAAATCTGCTCCGACATCAAGTAAAGGATTTG 1527
Db      2042 AAAGCTCTCAACGAAATGTGGAAGTGTCAGAACATGCTTCGGAGCCATGTACGCGAACTA 2101

Qy      1528 CTTGACTTGATTAAGCAACCCAAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGTG 1587
Db      2102 TTGGATTTGCACAAGCAGCCTACATCAGAGGCTAACTGTTCTGCCATGTTTGGAAAACTG 2161

Qy      1588 ATGGTTATTACAAGAAATTTACCTGATCCTGGTAAGGCTCAGGATTTCATGAAGAAATTC 1647
Db      2162 ATGACCATAGCAAAGAATTTGCCTGACCCCGGGAAAGCACAAGATTTTGTGAAGAAATTT 2221

Qy      1648 ACACAGGTGTTAGAAGATGATGAGAAAATAAGAAAGCAGTTAGAAGTACTTGTTAGTCCA 1707
Db      2222 AACCAGGTTCTCGGCGATGATGAGAAACTTCGGTCTCAGTTGGAGTTATTAATTAGCCCA 2281

Qy      1708 ACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1740
Db      2282 ACCTGTTCTTGCAAACAAGCAGATATTTGTGTG 2314
```

```
RESULT 13
AK026889
LOCUS       AK026889                2079 bp    mRNA    linear   PRI 29-SEP-2000
DEFINITION  Homo sapiens cDNA: FLJ23236 fis, clone COL00725.
ACCESSION   AK026889
VERSION     AK026889.1  GI:10439854
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
            Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
            Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2079)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Laboratory of Genome Structure Analysis, Human
            Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
            Japan (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Department of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1. .2079
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="COL00725"
                     /tissue_type="colon"
                     /clone_lib="COL"
                     /note="cloning vector pME18SFL3"
     CDS             304. .681
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB15584.1"
                     /db_xref="GI:10439855"
                     /translation="MDDLTKLVQEQKPKGSQRSRKRGHTASESDEQQWPEEKRLKEDI
```

```
                  LENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEER
                  QSGNTEQKSKSKQHRVSRRAQQR"
BASE COUNT      759 a    354 c    339 g    627 t
ORIGIN

  Query Match             10.7%;  Score 447;  DB 9;  Length 2079;
  Best Local Similarity   85.5%;  Pred. No. 1.3e-81;
  Matches  550;  Conservative    0;  Mismatches    0;  Indels   93;  Gaps    1;

Qy       3624 GTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACGCCCGTCACAGAACAGGAGGAGAA 3683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        237 GTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACGCCCGTCACAGAACAGGAGGAGAA 296

Qy       3684 ATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAACAGAAACCTAAAGGCAGTCAGCG 3743
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        297 ATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAACAGAAACCTAAAGGCAGTCAGCG 356

Qy       3744 AAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGTGGCCTGAGGAAAA 3803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        357 AAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGTGGCCTGAGGAAAA 416

Qy       3804 GAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCAAAAAAGGG 3863
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        417 GAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCAAAAAAGGG 476

Qy       3864 TAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAGCCAACAAT 3923
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        477 TAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAGCCAACAAT 536

Qy       3924 GAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGGAGGA 3983
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        537 GAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGGAGGA 596

Qy       3984 AGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAACAGCACCGAGT 4043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        597 AGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAACAGCACCGAGT 656

Qy       4044 GTCAAGGAGAGCACAGCAG----------------------------------------- 4062
              |||||||||||||||||||
Db        657 GTCAAGGAGAGCACAGCAGAGGTAAGCATGTGTAACTCTAAACTGCATCTGTTTCGTTAC 716

Qy       4063 ----------------------------------------------------AGAGCAGA 4070
                                                                  ||||||||
Db        717 TATATTATAAATCATAATTTGATGCTATCCACATTTGGGTCTTCCCCAAAGCAGAGCAGA 776

Qy       4071 ATCTCCTGAATCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAG 4130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        777 ATCTCCTGAATCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAG 836

Qy       4131 ACCATCAAAAACGCCATCACCATCACAACCAAAAAAAAAATGTG 4173
              |||||||||||||||||||||||||||||||||||||||||||
Db        837 ACCATCAAAAACGCCATCACCATCACAACCAAAAAAAAAATGTG 879


RESULT 14
HS49J10
LOCUS       HS49J10               137246 bp    DNA     linear   PRI 12-OCT-2001
DEFINITION  Human DNA sequence from clone RP1-49J10 on chromosome 13. Contains
            the 3' end of the gene for androgen-induced prostate proliferative
            shutoff associated protein (KIAA0979), ESTs, STSs and GSSs,
            complete sequence.
ACCESSION   Z84572
VERSION     Z84572.1  GI:1813974
KEYWORDS    HTG; KIAA0979.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 137246)
  AUTHORS   Whiteley,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP1-49J10 is from the library RPCI-1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            IMPORTANT: This sequence is not the entire insert of clone
            RP1-49J10 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true right end of clone RP1-267P19 is at 99 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .137246
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP1-49J10"
                     /clone_lib="RPCI-1"
     repeat_region   520. .645
                     /note="L2 repeat: matches 2575. .2706 of consensus"
     repeat_region   1551. .1831
                     /note="AluSx repeat: matches 1. .281 of consensus"
     repeat_region   1838. .1873
                     /note="18 copies 2 mer tg 100% conserved"
     repeat_region   1876. .2848
                     /note="L1PA15 repeat: matches 5151. .6157 of consensus"
     repeat_region   3149. .3405
                     /note="MER44A repeat: matches 1. .247 of consensus"
     repeat_region   3406. .3530
                     /note="L2 repeat: matches 2556. .2704 of consensus"
     repeat_region   3815. .3852
                     /note="7SLRNA repeat: matches 279. .316 of consensus"
     repeat_region   3853. .3922
                     /note="AluSc repeat: matches 120. .188 of consensus"
     gene            4320. .26902
                     /gene="49J10.1"
     mRNA            join(<4320. .4524,6830. .6944,9075. .9207,9522. .9641,
                     10616. .10678,11563. .11708,15477. .15582,21108. .21547,
                     21638. .21748,24176. .24311,26004. .26902)
                     /gene="49J10.1"
                     /product="49J10.1.1 (KIAA0979, isoform 1)"
                     /note="match: cDNAs: Em:AB023196
                     match: ESTs: Em:BF043498 Em:BG571128 Em:BF094231 Em:F12354
                     Em:HSC38H111 Em:T64867 Em:BF104916 Em:BE891040 Em:H44781"
                     /evidence=not_experimental
     mRNA            join(<4320. .4524,6830. .6944,9075. .9207,9522. .9641,
                     10616. .10678,11563. .11708,15477. .15582,21108. .21547,
                     21641. .21865,26004. .26902)
                     /gene="49J10.1"
                     /product="49J10.1.2 (androgen-induced prostate
                     proliferative shutoff associated protein, isoform 2)"
                     /note="match: cDNAs: Em:U95825"
```

```
                 /evidence=not_experimental
CDS              join(<4320. .4524,6830. .6944,9075. .9207,9522. .9641,
                 10616. .10678,11563. .11708,15477. .15582,21108. .21547,
                 21638. .21748,24176. .24311,26004. .26039)
                 /gene="49J10.1"
                 /note="continued from 267P19.1 in Em:Z75889
                 match: proteins: Tr:Q9Y2I5"
                 /codon_start=1
                 /evidence=not_experimental
                 /product="49J10.1.1 (KIAA0979, isoform 1)"
                 /protein_id="CAC94788.1"
                 /db_xref="GI:16151354"
                 /translation="DECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRA
                 HARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQL
                 KDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVA
                 MNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNV
                 LGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYT
                 MSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGS
                 QRSRKRGHTASESDEQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKE
                 EPTMKTSKKGSKKKSGPPAPEEEEEEEERQSGNTEQKSKSKQHRVSRRAQQSRAESPES
                 SAIESTQSTPQKGRGRPSKTPSPSQPKKNVRVGRSKQAATKENDSSEEVDVFQGSSPV
                 DDIPQEETEEEEVSTVNVRRRSAKRERR"
CDS              join(<4320. .4524,6830. .6944,9075. .9207,9522. .9641,
                 10616. .10678,11563. .11708,15477. .15582,21108. .21547,
                 21641. .21752)
                 /gene="49J10.1"
                 /note="continued from 267P19.1 in Em:Z75889
                 match: proteins: Tr:Q9Y451"
                 /codon_start=1
                 /evidence=not_experimental
                 /product="49J10.1.2 (androgen-induced prostate
                 proliferative shutoff associated protein, isoform 2)"
                 /protein_id="CAC94787.1"
                 /db_xref="GI:16151353"
                 /translation="DECYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRA
                 HARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQL
                 KDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVA
                 MNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNV
                 LGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYT
                 MSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGS
                 QRSRKRGHTASESDEQQWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKE
                 EPTMKTSKKGSKKKSGPPAPEEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESPESS
                 AIESTQSTPQKGRGRPSKTPSPSQPKKNV"
repeat_region    4959. .5316
                 /note="L2 repeat: matches 1830. .2229 of consensus"
repeat_region    5608. .5810
                 /note="MIR repeat: matches 46. .262 of consensus"
repeat_region    5874. .5953
                 /note="L2 repeat: matches 2625. .2736 of consensus"
repeat_region    6757. .6826
                 /note="35 copies 2 mer tt 80% conserved"
repeat_region    7829. .8133
                 /note="AluSx repeat: matches 1. .297 of consensus"
repeat_region    8892. .9087
                 /note="98 copies 2 mer tt 59% conserved"
repeat_region    10957. .11242
                 /note="AluY repeat: matches 1. .291 of consensus"
repeat_region    12344. .13119
                 /note="L1PA16 repeat: matches 5321. .6156 of consensus"
repeat_region    14108. .14427
                 /note="L2 repeat: matches 2014. .2333 of consensus"
repeat_region    14428. .14540
                 /note="HY1 repeat: matches 1. .111 of consensus"
repeat_region    14541. .14935
                 /note="L2 repeat: matches 1619. .2014 of consensus"
repeat_region    14951. .15154
                 /note="L2 repeat: matches 2529. .2730 of consensus"
repeat_region    16579. .16885
                 /note="AluSx repeat: matches 4. .312 of consensus"
repeat_region    18059. .18397
                 /note="L1P repeat: matches 3809. .4138 of consensus"
repeat_region    19076. .19481
                 /note="MLT1A2 repeat: matches 1. .374 of consensus"
repeat_region    19502. .19807
                 /note="AluSx repeat: matches 1. .310 of consensus"
repeat_region    19807. .20122
                 /note="AluY repeat: matches 1. .294 of consensus"
repeat_region    23282. .23371
                 /note="L2 repeat: matches 2590. .2733 of consensus"
repeat_region    23734. .23944
                 /note="L2 repeat: matches 1814. .2050 of consensus"
repeat_region    24795. .25089
                 /note="AluSx repeat: matches 1. .294 of consensus"
repeat_region    26083. .26106
                 /note="12 copies 2 mer tt 95% conserved"
misc_feature     26140. .26289
                 /gene="49J10.1"
                 /note="match: STS: Em:G43328 Em:G43329"
misc_feature     26749. .27206
                 /note="match: GSS: Em:AZ019111"
misc_feature     26790. .27245
                 /note="match: GSS: Em:AQ985541"
polyA_signal     26877. .26882
                 /gene="49J10.1"
polyA_site       26899
                 /gene="49J10.1"
polyA_site       26900
                 /gene="49J10.1"
polyA_site       26901
                 /gene="49J10.1"
polyA_site       26902
                 /gene="49J10.1"
repeat_region    28327. .28360
                 /note="17 copies 2 mer tt 88% conserved"
misc_feature     29194. .29299
                 /note="match: GSS: Em:AQ675002"
misc_feature     29212. .29650
                 /note="match: GSS: Em:AQ675002"
repeat_region    29300. .29554
                 /note="L2 repeat: matches 2498. .2750 of consensus"

Query Match             10.7%;  Score 447;  DB 9;  Length 137246;
Best Local Similarity   85.5%;  Pred. No. 6e-82;
Matches  550;  Conservative    0;  Mismatches    0;  Indels   93;  Gaps    1;

Qy       3624 GTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACGCCCGTCACAGAACAGGAGGAGAA 3683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21107 GTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACGCCCGTCACAGAACAGGAGGAGAA 21166

Qy       3684 ATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAACAGAAACCTAAAGGCAGTCAGCG 3743
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21167 ATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAACAGAAACCTAAAGGCAGTCAGCG 21226

Qy       3744 AAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGTGGCCTGAGGAAAA 3803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21227 AAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGTGGCCTGAGGAAAA 21286

Qy       3804 GAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCAAAAAAGGG 3863
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21287 GAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCAAAAAAGGG 21346

Qy       3864 TAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAGCCAACAAT 3923
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21347 TAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAGCCAACAAT 21406

Qy       3924 GAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGGAGGA 3983
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21407 GAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGGAGGA 21466

Qy       3984 AGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAACAGCACCGAGT 4043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21467 AGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAACAGCACCGAGT 21526

Qy       4044 GTCAAGGAGAGCACAGCAG----------------------------------------- 4062
              |||||||||||||||||||
Db      21527 GTCAAGGAGAGCACAGCAGAGGTAAGCATGTGTAACTCTAAACTGCATCTGTTTCGTTAC 21586
```

```
Qy       4063 ----------------------------------------------------AGAGCAGA 4070
                                                                  ||||||||
Db      21587 TATATTATAAATCATAATTTGATGCTATCCACATTTGGGTCTTCCCCAAAGCAGAGCAGA 21646

Qy       4071 ATCTCCTGAATCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAG 4130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21647 ATCTCCTGAATCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAG 21706

Qy       4131 ACCATCAAAAACGCCATCACCATCACAACCAAAAAAAAATGTG 4173
              |||||||||||||||||||||||||||||||||||||||||||
Db      21707 ACCATCAAAAACGCCATCACCATCACAACCAAAAAAAAATGTG 21749


RESULT 15
AL138820
LOCUS       AL138820               168487 bp    DNA     linear   PRI 01-NOV-2000
DEFINITION  Human DNA sequence from clone RP11-380B4 on chromosome 13, complete
            sequence.
ACCESSION   AL138820
VERSION     AL138820.11  GI:11022364
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168487)
  AUTHORS   Smith,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 25, 2000 this sequence version replaced gi:10715762.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-380B4 is from the library RPCI-11.2 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-380B4.
FEATURES             Location/Qualifiers
     source          1. .168487
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-380B4"
                     /clone_lib="RPCI-11.2"
BASE COUNT    48433 a   33155 c   34624 g   52275 t
ORIGIN

  Query Match             10.7%;  Score 447;  DB 9;  Length 168487;
  Best Local Similarity   85.5%;  Pred. No. 5.7e-82;
  Matches  550;  Conservative    0;  Mismatches    0;  Indels   93;  Gaps    1;

Qy       3624 GTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACGCCCGTCACAGAACAGGAGGAGAA 3683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20593 GTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACGCCCGTCACAGAACAGGAGGAGAA 20652

Qy       3684 ATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAACAGAAACCTAAAGGCAGTCAGCG 3743
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20653 ATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAACAGAAACCTAAAGGCAGTCAGCG 20712

Qy       3744 AAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGTGGCCTGAGGAAAA 3803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20713 AAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGATGAACAGCAGTGGCCTGAGGAAAA 20772

Qy       3804 GAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCAAAAAAGGG 3863
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20773 GAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAACAGAATAGTCCGCCAAAAAAGGG 20832

Qy       3864 TAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAGCCAACAAT 3923
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20833 TAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGTACACCAAAAGAAGAGCCAACAAT 20892

Qy       3924 GAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGGAGGA 3983
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20893 GAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGACCTCCAGCACCAGAGGAGGAGGA 20952

Qy       3984 AGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAACAGCACCGAGT 4043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20953 AGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAGTCCAAAAGCAAACAGCACCGAGT 21012

Qy       4044 GTCAAGGAGAGCACAGCAG----------------------------------------- 4062
              |||||||||||||||||||
Db      21013 GTCAAGGAGAGCACAGCAGAGGTAAGCATGTGTAACTCTAAACTGCATCTGTTTCGTTAC 21072

Qy       4063 ----------------------------------------------------AGAGCAGA 4070
                                                                  ||||||||
Db      21073 TATATTATAAATCATAATTTGATGCTATCCACATTTGGGTCTTCCCCAAAGCAGAGCAGA 21132

Qy       4071 ATCTCCTGAATCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAG 4130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21133 ATCTCCTGAATCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAG 21192

Qy       4131 ACCATCAAAAACGCCATCACCATCACAACCAAAAAAAAATGTG 4173
              |||||||||||||||||||||||||||||||||||||||||||
Db      21193 ACCATCAAAAACGCCATCACCATCACAACCAAAAAAAAATGTG 21235


Search completed: September 24, 2003, 14:13:31
Job time : 14881 secs
```

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         September 24, 2003, 20:22:05 ; Search time 11608 Seconds
                                          (without alignments)
                                          4902.248 Million cell updates/sec

Title:          US-09-512-581B-2
Perfect score:  7193
Sequence:       1 MAHSKTRTNDGKITYPPGVK..........QKGRGRPSKTPSPSQPKKNV 1391

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters:      5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09512581/runat_23092003_163546_5474/app_query.fasta_1.1543
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09512581_@CGN_1_1_7890_@runat_23092003_163546_5474 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*
                15:  em_ba:*
                16:  em_fun:*
                17:  em_hum:*
                18:  em_in:*
                19:  em_mu:*
                20:  em_om:*
                21:  em_or:*
                22:  em_ov:*
                23:  em_pat:*
                24:  em_ph:*
                25:  em_pl:*
                26:  em_ro:*
                27:  em_sts:*
                28:  em_un:*
                29:  em_vi:*
                30:  em_htg_hum:*
                31:  em_htg_inv:*
                32:  em_htg_other:*
                33:  em_htg_mus:*
                34:  em_htg_pln:*
                35:  em_htg_rod:*
                36:  em_htg_mam:*
                37:  em_htg_vrt:*
                38:  em_sy:*
                39:  em_htgo_hum:*
                40:  em_htgo_mus:*
                41:  em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	7155	99.5	5309	9	AB023196	AB023196 Homo sapi
	2	7155	99.5	7444	9	HS26H231	AL137201 Novel hum
	3	7101	98.7	5253	9	HSU95825	U95825 Human andro
	4	6942	96.5	5317	10	AY102267	AY102267 Mus muscu
	5	5659	78.7	5944	10	AK122414	AK122414 Mus muscu
	6	4760.5	66.2	6744	9	AF294791	AF294791 Homo sapi
	7	2953.5	41.1	1852	9	HSU50533	U50533 Human BRCA2
	8	2704.5	37.6	5177	9	AB014548	AB014548 Homo sapi
	9	2517	35.0	4252	9	BC039256	BC039256 Homo sapi
	10	2474	34.4	2265	9	AK098331	AK098331 Homo sapi
	11	2474	34.4	2705	9	BC041361	BC041361 Homo sapi
	12	2442	33.9	1976	10	BC021408	BC021408 Mus muscu
	13	1836	25.5	216608	5	AL844559	AL844559 Zebrafish
	14	1729.5	24.0	173613	3	AC007475	AC007475 Drosophil
	15	1729.5	24.0	192763	3	AC007474	AC007474 Drosophil
	16	1729.5	24.0	292919	3	AE003823	AE003823 Drosophil
c	17	1708.5	23.8	194634	2	AC020286	AC020286 Drosophil
	18	925.5	12.9	2079	9	AK026889	AK026889 Homo sapi
	19	925.5	12.9	137246	9	HS49J10	Z84572 Human DNA s
	20	925.5	12.9	168487	9	AL138820	AL138820 Human DNA
	21	840.5	11.7	3779	9	BC009650	BC009650 Homo sapi
	22	833.5	11.6	4046	8	AB067651	AB067651 Schizosac
	23	833.5	11.6	9982	8	SPAC110	AL441624 S.pombe c
c	24	819.5	11.4	54398	2	AC068224	AC068224 Homo sapi
	25	769	10.7	236508	2	AC111126	AC111126 Mus muscu
c	26	768	10.7	72157	2	AC016449	AC016449 Homo sapi
c	27	747	10.4	263479	2	AC125986	AC125986 Rattus no
	28	706.5	9.8	2212	6	BD159124	BD159124 Primer fo
	29	706.5	9.8	2212	9	AK021757	AK021757 Homo sapi
	30	698	9.7	439	6	BD059747	BD059747 Secreted
	31	697.5	9.7	6411	8	SMAJ9934	AJ009934 Sordaria
	32	645	9.0	772	6	BD145886	BD145886 Primer fo
	33	635	8.8	4885	8	EMEBIMD	L03200 Emericella
	34	584	8.1	153751	3	AC116551	AC116551 Dictyoste
	35	554.5	7.7	54398	2	AC068224	AC068224 Homo sapi
c	36	523	7.3	184321	2	AC110493	AC110493 Mus muscu
c	37	523	7.3	190058	2	AC112263	AC112263 Mus muscu
c	38	523	7.3	298053	2	AC116574	AC116574 Mus muscu
c	39	520	7.2	233589	2	AC095342	AC095342 Rattus no
	40	512	7.1	163355	2	AC117927	AC117927 Mus muscu
	41	505	7.0	1541	3	AY058358	AY058358 Drosophil
	42	501	7.0	295	6	BD026562	BD026562 Sequence
c	43	466	6.5	98856	9	AC113150	AC113150 Homo sapi
c	44	466	6.5	145450	2	AC068352	AC068352 Homo sapi
	45	401.5	5.6	85592	2	AL512630	AL512630 Mus muscu

ALIGNMENTS

RESULT 1

```
AB023196
LOCUS       AB023196                5309 bp    mRNA    linear   PRI 10-MAY-2002
DEFINITION  Homo sapiens mRNA for KIAA0979 protein, partial cds.
ACCESSION   AB023196
VERSION     AB023196.2  GI:20521717
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirosawa,M.,
            Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            XIII. The complete sequences of 100 new cDNA clones from brain
            which code for large proteins in vitro
  JOURNAL   DNA Res. 6 (1), 63-70 (1999)
  MEDLINE   99246063
   PUBMED   10231032
REFERENCE   2  (bases 1 to 5309)
  AUTHORS   Ohara,O., Nagase,T. and Kikuno,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-1999) Osamu Ohara, Kazusa DNA Research Institute,
            Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
            292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913,
            Fax:+81-438-52-3914)
COMMENT     On May 9, 2002 this sequence version replaced gi:4589601.
FEATURES             Location/Qualifiers
     source          1..5309
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="hj07056s1"
                     /sex="male"
                     /tissue_type="brain"
                     /clone_lib="pBluescriptII SK plus"
                     /dev_stage="adult"
                     /note="This sequence was obtained bysubcloning of the DNA
                     fragments derived from two cDNA clones (1 - 722 was
                     derived from fj09522 and 723 - 5309 was derived from
                     hj07056)."
     gene            1..5309
                     /gene="KIAA0979"
     CDS             <1..4452
                     /gene="KIAA0979"
                     /codon_start=1
                     /product="KIAA0979 protein"
                     /protein_id="BAA76823.2"
                     /db_xref="GI:20521718"
                     /translation="GGGRRRLDAAAAADPGEPRSERSSESATRRGRNISVMAHSKTRT
                     NDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMDMDQDSEEEKELYLNLALHLASD
                     FFLKHPDKDVRLLVACCLADIFRIYAPEAPYTSPDKLKDIFMFITRQLKGLEDTKSPQ
                     FNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRTLFSVINNGHNQKVHMHMVDLMS
                     SIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLAKALLKRTAQAIEPYITNFFNQV
                     LMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAK
                     MFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLK
                     VRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQ
                     IYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTE
                     RMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVM
                     VITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVLVSPTCSCKQAEGCVREITKKLG
                     NPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAI
                     RAGLELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFP
                     HIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEH
                     LITPLVTIGHIALLAPDQFAAPLKSLVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSP
                     ETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRL
                     RLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLP
                     LEYMAICALCAKDPVKERRAHARQCLVKNINVRREYLKQHAAVSEKLLSLLPEYVVPY
                     TIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKD
                     AQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKN
                     YLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPG
                     RIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKRDDSDLVRSELEKPRGRKKTPVTEQE
                     EKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESDEQQWPEEKRLKEDILENEDEQNSP
                     PKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSGPPAPEEEEEEERQSGNTEQKSK
                     SKQHRVSRRAQQRAESPESSAIESTQSTPQKGRGRPSKTPSPSQPKKNVRVGRSKQAA
                     TKENDSSEEVDVFQGSSPVDDIPQEETEEEEVSTVNVRRRSAKRERR"
BASE COUNT     1774 a    955 c   1115 g   1465 t
ORIGIN

Alignment Scores:
Pred. No.:              0            Length:         5309
Score:                  7155.00      Matches:        1387
Percent Similarity:     99.71%       Conservative:   0
Best Local Similarity:  99.71%       Mismatches:     4
Query Match:            99.47%       Indels:         0
DB:                     9            Gaps:           0

US-09-512-581B-2 (1-1391) x AB023196 (1-5309)

Qy          1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProProGlyValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        109 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 168

Qy         21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValValLys 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        169 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 228

Qy         41 ThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        229 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 288

Qy         61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
              |||||||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db        289 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGATAAAGATGTTCGCTTACTGGTA 348

Qy         81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        349 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 408

Qy        101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        409 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 468

Qy        121 LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        469 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 528

Qy        141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        529 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 588

Qy        161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        589 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 648

Qy        181 MetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        649 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 708

Qy        201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        709 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 768

Qy        221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
              |||||||||||||||||||||||||||||||||||||||||    |||||||||||||||
Db        769 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTTTTAATCAGGTT 828

Qy        241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        829 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 888

Qy        261 GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        889 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 948

Qy        281 LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      949 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 1008

Qy      301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGly 320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1009 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 1068

Qy      321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1069 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1128

Qy      341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1129 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1188

Qy      361 AspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys 380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1189 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1248

Qy      381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1249 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1308

Qy      401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAla 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1309 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1368

Qy      421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu 440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1369 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1428

Qy      441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1429 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1488

Qy      461 GlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1489 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1548

Qy      481 TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1549 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1608

Qy      501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1609 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1668

Qy      521 SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1669 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1728

Qy      541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArg 560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1729 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1788

Qy      561 LysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysVal 580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1789 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1848

Qy      581 ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1849 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1908

Qy      601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla 620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1909 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 1968

Qy      621 LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyVal 640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1969 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 2028

Qy      641 ProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHis 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2029 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 2088

Qy      661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2089 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2148

Qy      681 AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2149 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2208

Qy      701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLys 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2209 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2268

Qy      721 GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys 740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2269 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2328

Qy      741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2329 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2388

Qy      761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2389 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2448

Qy      781 PheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
            |||||||||||    ||||||    |||||||||||||||||||||||||||||||||||
Db     2449 TTTGCTGCTCCTTTGAAATCTTTGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2508

Qy      801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerPro 820
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2509 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2568

Qy      821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLys 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2569 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2628

Qy      841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2629 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2688

Qy      861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2689 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2748

Qy      881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2749 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2808

Qy      901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2809 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2868

Qy      921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet 940
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2869 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2928

Qy      941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGln 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2929 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 2988

Qy      961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSer 980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2989 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 3048

Qy      981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3049 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3108

Qy     1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3109 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3168
```

```
Qy       1021 LeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle 1040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3169 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3228

Qy       1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3229 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3288

Qy       1061 MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3289 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3348

Qy       1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln 1100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3349 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3408

Qy       1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe 1120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3409 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3468

Qy       1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3469 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3528

Qy       1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3529 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3588

Qy       1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3589 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3648

Qy       1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3649 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3708

Qy       1201 ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThr 1220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3709 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3768

Qy       1221 ProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlu 1240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3769 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3828

Qy       1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3829 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3888

Qy       1261 GluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGlu 1280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3889 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3948

Qy       1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3949 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 4008

Qy       1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLysSerGly 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4009 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 4068

Qy       1321 ProProAlaProGluGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys 1340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4069 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4128

Qy       1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4129 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4188

Qy       1361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4189 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4248

Qy       1381 ThrProSerProSerGlnProLysLysAsnVal 1391
              |||||||||||||||||||||||||||||||||
Db       4249 ACGCCATCACCATCACAACCAAAAAAAAATGTC 4281
```

NPA

```
RESULT 2
HS26H231
LOCUS       HS26H231                7444 bp    mRNA    linear   PRI 25-JAN-2000
DEFINITION  Novel human gene mapping to chomosome 13.
ACCESSION   AL137201
VERSION     AL137201.1  GI:6759511
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 7444)
  AUTHORS   Rhodes,S. and Huckle,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-2000) E-mail contact: humquery@sanger.ac.uk
COMMENT     This cDNA sequence was assembled from public domain ESTs and single
            pass sequencing reads from expressed DNA templates, aligned to the
            genomic DNA sequence from the bacterial clones 26H23 (Z84467),
            267P19 (Z75889) and 49J10( Z84572).
            The EST sequences listed match this sequence with an identity of at
            least 95% between the coordinates shown.
            Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13/ Experimentally determined gene
            Sanger Centre name : 26H23.C13.1.
FEATURES             Location/Qualifiers
     source          1. .7444
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /map="13q12.3"
     exon            1. .107
                     /number=1
     exon            108. .234
                     /number=2
     CDS             127. .4470
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB69911.1"
                     /db_xref="GI:6759512"
                     /translation="MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMD
                     MDQDSEEEKELYLNLALHLASDFFLKHPDKDVRLLVACCLADIFRIYAPEAPYTSPDK
                     LKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRT
                     LFSVINNGHNQKVHMHMVDLMSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLA
                     KALLKRTAQAIEPYITNFFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQ
                     LEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECV
                     KFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNF
                     VRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSID
                     DRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKD
                     LLDLIKQPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVL
                     VSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQ
                     VNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDE
                     KVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSK
                     ETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPLKSLVATFIVKDLL
                     MNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTT
                     ILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDE
                     CYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREY
                     LKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMA
                     KNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESP
                     KDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQ
                     TKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKRDDS
                     DLVRSELEKPRGRKKTPVTEQEEKLGMDDLTKLVQEQKPKGSQRSRKRGHTASESDEQ
                     QWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSG
                     PPAPEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESPESSAIESTQSTPQKGRGRP
                     SKTPSPSQPKKNVRVGRSKQAATKENDSSEEVDVFQGSSPVDDIPQEETEEEEVSTVN
                     VRRRSAKRERR"
     exon            235. .438
                     /number=3
     exon            439. .525
                     /number=4
```

```
     exon            526. .623
                     /number=5
     exon            624. .750
                     /number=6
     exon            751. .831
                     /number=7
     exon            832. .972
                     /number=8
     exon            973. .1088
                     /number=9
     exon            1089. .1183
                     /number=10
     exon            1184. .1329
                     /number=11
     exon            1330. .1481
                     /number=12
     exon            1482. .1595
                     /number=13
     exon            1596. .1677
                     /number=14
     exon            1678. .1726
                     /number=15
     exon            1727. .1866
                     /number=16
     misc_feature    1866. .2046
                     /note="matches EST AA687846 from clone IMAGE:1219676"
     misc_feature    1866. .1941
                     /note="matches EST AA687776 from clone IMAGE:1219628"
     exon            1867. .1982
                     /number=17
     exon            1983. .2088
                     /number=18
     exon            2089. .2249
                     /number=19
     exon            2250. .2373
                     /number=20
     exon            2374. .2532
                     /number=21
     exon            2533. .2601
                     /number=22
     exon            2602. .2738
                     /number=23
     exon            2739. .2862
                     /number=24
     exon            2863. .3067
                     /number=25
     misc_feature    complement(2912. .3302)
                     /note="matches EST AA630616 from clone 856164"
     exon            3068. .3182
                     /number=26
     exon            3183. .3315
                     /number=27
     misc_feature    3302. .3572
                     /note="matches EST AA333118"
     exon            3316. .3435
                     /number=28
     misc_feature    3349. .3818
                     /note="matches EST AA282190 from clone IMAGE:704631"
     exon            3436. .3498
                     /number=29
     exon            3499. .3644
                     /number=30
     misc_feature    complement(3586. .4065)
                     /note="matches EST AI655429 from clone IMAGE:2239331"
     exon            3645. .3750
                     /number=31
     exon            3751. .4190
                     /number=32
     misc_feature    complement(3864. .4286)
                     /note="matches EST AI680124 from clone IMAGE:2264521"
     misc_feature    4001. .4119
                     /note="matches EST N83765 from clone KK2745"
     misc_feature    4135. .4450
                     /note="matches EST F12354 from clone c-38h11."
     exon            4191. .4298
                     /number=33
     misc_feature    4255. .4307
                     /note="matches EST T64867 from clone 22264"
     exon            4299. .4434
                     /number=34
     misc_feature    join(4398. .4533,4518. .4660,4659. .4675,4674. .4693)
                     /note="matches EST H44781 from clone 188032"
     exon            4435. .7444
                     /number=35
     misc_feature    4537. .5022
                     /note="matches EST AI681422 from clone IMAGE:2272632"
     misc_feature    4537. .4892
                     /note="matches EST AA911125 from clone IMAGE:1416862"
     misc_feature    4538. .4969
                     /note="matches EST AI275866 from clone IMAGE:1877482"
     misc_feature    4538. .4947
                     /note="matches EST AA987361 from clone IMAGE:1603350"
     misc_feature    4538. .4943
                     /note="matches EST AI299646 from clone IMAGE:1898010"
     misc_feature    4562. .5003
                     /note="matches EST AA447433 from clone 784555"
     misc_feature    4571. .4854
                     /note="matches EST H93424 from clone 220700"
     misc_feature    join(4573. .4649,4650. .4815)
                     /note="matches EST W20070 from clone 306049"
     misc_feature    complement(4629. .4937)
                     /note="matches EST AW089775 from clone IMAGE:2593146"
     misc_feature    4746. .4994
                     /note="matches EST AA747568 from clone IMAGE:1269696"
     misc_feature    complement(4803. .5332)
                     /note="matches EST AI911784 from clone IMAGE:2329873"
     misc_feature    complement(4828. .5296)
                     /note="matches EST AA563884 from clone IMAGE:1013990"
     misc_feature    4861. .5243
                     /note="matches EST D20082 from clone pm1283."
     misc_feature    join(4872. .4921,5281. .5371)
                     /note="matches EST AA177347 from clone 621810"
     misc_feature    complement(join(4872. .4921,5281. .5332))
                     /note="matches EST AW212954 from clone IMAGE:2647801"
     misc_feature    4872. .4921
                     /note="matches EST AA154932 from clone 599194"
     misc_feature    complement(4872. .4921)
                     /note="matches EST AW048929 from clone
                     UI-M-BH1-amo-c-09-0-UI"
     misc_feature    complement(4876. .5333)
                     /note="matches EST AI016896 from clone IMAGE:1627886"
     misc_feature    complement(join(4922. .4975,5281. .5332))
                     /note="matches EST AI012649 from clone RPLBA02"
     misc_feature    complement(4922. .4975)
                     /note="matches EST AA899586 from clone
                     UI-R-E0-cy-e-02-0-UI
                     matches EST AI172172 from clone RMUBV82"
     misc_feature    complement(4931. .5332)
                     /note="matches EST AA625961 from clone 745478"
     misc_feature    complement(4951. .5332)
                     /note="matches EST AI140790 from clone IMAGE:1566222"
     misc_feature    complement(5021. .5309)

Alignment Scores:
Pred. No.:                0             Length:         7444
Score:                    7155.00       Matches:        1387
Percent Similarity:       99.71%        Conservative:   0
Best Local Similarity:    99.71%        Mismatches:     4
Query Match:              99.47%        Indels:         0
DB:                       9             Gaps:           0

US-09-512-581B-2 (1-1391) x HS26H231 (1-7444)

Qy          1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProProGlyValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 186
```

```
Qy         21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValValLys 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 246

Qy         41 ThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 306

Qy         61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
              ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db        307 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGATAAAGATGTTCGCTTACTGGTA 366

Qy         81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 426

Qy        101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 486

Qy        121 LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        487 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 546

Qy        141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        547 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 606

Qy        161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        607 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 666

Qy        181 MetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        667 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 726

Qy        201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        727 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 786

Qy        221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
              ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db        787 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTTTTAATCAGGTT 846

Qy        241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        847 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 906

Qy        261 GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        907 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 966

Qy        281 LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        967 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 1026

Qy        301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGly 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1027 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 1086

Qy        321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1087 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1146

Qy        341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1147 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1206

Qy        361 AspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1207 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1266

Qy        381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1267 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1326

Qy        401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAla 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1327 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1386

Qy        421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1387 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1446

Qy        441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1447 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1506

Qy        461 GlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1507 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1566

Qy        481 TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1567 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1626

Qy        501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1627 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1686

Qy        521 SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1687 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAATTTACCTGATCCTGGT 1746

Qy        541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArg 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1747 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1806

Qy        561 LysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysVal 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1807 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1866

Qy        581 ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1867 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1926

Qy        601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1927 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 1986

Qy        621 LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyVal 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1987 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 2046

Qy        641 ProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHis 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2047 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 2106

Qy        661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2107 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCTTGTCTGAAAATGGAT 2166

Qy        681 AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2167 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2226

Qy        701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLys 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2227 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2286

Qy        721 GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2287 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2346

Qy        741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2347 GAGACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2406

Qy          761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2407 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2466

Qy          781 PheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
                |||||||||||   ||||||   |||||||||||||||||||||||||||||||||||||
Db         2467 TTTGCTGCTCCTTTGAAATCTTTGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2526

Qy          801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerPro 820
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2527 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2586

Qy          821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLys 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2587 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2646

Qy          841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2647 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2706

Qy          861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2707 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2766

Qy          881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2767 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2826

Qy          901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2827 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2886

Qy          921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet 940
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2887 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2946

Qy          941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGln 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2947 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 3006

Qy          961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSer 980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3007 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 3066

Qy          981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3067 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3126

Qy         1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3127 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3186

Qy         1021 LeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle 1040
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3187 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3246

Qy         1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3247 AGAAAGATGGTAGAAAATATTAAACAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3306

Qy         1061 MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3307 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3366

Qy         1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln 1100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3367 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3426

Qy         1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe 1120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3427 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGAAATGAAATCATTTTTC 3486

Qy         1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3487 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3546

Qy         1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3547 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAAGCAATGCAAGCAGC 3606

Qy         1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3607 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3666

Qy         1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3667 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAAAAGTGACAAG 3726

Qy         1201 ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThr 1220
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3727 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGGAAAAAAACG 3786

Qy         1221 ProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlu 1240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3787 CCCGTCACAGAACAGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3846

Qy         1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp 1260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3847 CAGAAACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3906

Qy         1261 GluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGlu 1280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3907 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3966

Qy         1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3967 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 4026

Qy         1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLysSerGly 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         4027 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 4086

Qy         1321 ProProAlaProGluGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys 1340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         4087 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4146

Qy         1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         4147 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTCCTGAA 4206

Qy         1361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         4207 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4266

Qy         1381 ThrProSerProSerGlnProLysLysAsnVal 1391
                |||||||||||||||||||||||||||||||||
Db         4267 ACGCCATCACCATCACAACCAAAAAAAAATGTC 4299
```

```
RESULT 3
HSU95825
LOCUS       HSU95825             5253 bp    mRNA    linear   PRI 05-APR-1999
DEFINITION  Human androgen-induced prostate proliferative shutoff associated
            protein (AS3) mRNA, complete cds.
ACCESSION   U95825
VERSION     U95825.2  GI:4559409
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5253)
  AUTHORS   Geck,P., Szelei,J., Jimenez,J., Soto,A.M. and Sonnenschein,C.
  TITLE     Androgen-induced proliferative shutoff in prostate cancer cells
```

```
  JOURNAL   Proc. Annu. Meet. Am. Assoc. Cancer Res. 37, 223-223 (1996)
REFERENCE   2  (bases 1 to 5253)
  AUTHORS   Geck,P., Szelei,J., Jimenez,J., Sonnenschein,C. and Soto,A.M.
  TITLE     Early gene expression during androgen-induced inhibition of
            proliferation of prostate cancer cells: a new suppressor candidate
            on chromosome 13, in the BRCA2-Rb1 locus
  JOURNAL   J. Steroid Biochem. Mol. Biol. 68 (1-2), 41-50 (1999)
  MEDLINE   99229875
   PUBMED   10215036
REFERENCE   3  (bases 1 to 5253)
  AUTHORS   Geck,P., Szelei,J., Jimenez,J., Sonnenschein,C. and Soto,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-1997) Anatomy and Cell Biology, Tufts University
            Medical School, 136 Harrison Avenue, Boston, MA 02111, USA
REFERENCE   4  (bases 1 to 5253)
  AUTHORS   Geck,P., Szelei,J., Jimenez,J., Sonnenschein,C. and Soto,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-1999) Anatomy and Cell Biology, Tufts University
            Medical School, 136 Harrison Avenue, Boston, MA 02111, USA
  REMARK    Sequence update by submitter
COMMENT     On Apr 5, 1999 this sequence version replaced gi:4539617.
FEATURES             Location/Qualifiers
     source          1..5253
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /map="13q12-q13"
                     /cell_line="LNCaP"
                     /cell_type="cancer cells"
                     /tissue_type="prostate"
                     /note="exon 1 maps to P1 Artificial Chromosome PAC26H23;
                     exons 2-24 map to cosmid 267p19; exons 25-34 map to P1
                     Artificial Chromosome PAC 49J10"
     gene            1..5253
                     /gene="AS3"
     exon            1..46
                     /gene="AS3"
                     /number=1
     exon            47..173
                     /gene="AS3"
                     /number=2
     CDS             66..4241
                     /gene="AS3"
                     /codon_start=1
                     /product="androgen-induced prostate proliferative shutoff
                     associated protein"
                     /protein_id="AAD22134.2"
                     /db_xref="GI:4559410"
                     /translation="MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMD
                     MDQDSEEEKELYLNLALHLASDFFLKHPDKDVRLLVACCLADIFRIYAPEAPYTSPDK
                     LKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRT
                     LFSVINNGHNQKVHMHMVDLMSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLA
                     KALLKRTAQAIEPYITNFFNQVLMLGKTSISDLSEHVSDLILELYNIDSHLLLSVLPQ
                     LEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECV
                     KFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNF
                     VGERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSID
                     DRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKD
                     LLDLIKQPKTDASVKAIFSKVMVITRSLPDPGKAQDFMKKFTQVLEDDEKIRKQLEVL
                     VSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQ
                     VNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDE
                     KVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSK
                     EAQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPLKSLVATFIVKDLL
                     MNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTT
                     ILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDE
                     CYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNINVRREY
                     LKQHAAVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMA
                     KNENNSHAFIRKMVENIKQTKDAQGPDDAKMNEKLYTVCDVAMNIIMSKSTTYSLESP
                     KDPVLPARFFTQPDKNFSNTKNYLPPGMKSFFTPGKPKTTNVLGAVNKPLSSAGKQSQ
                     TKSSRMETVGNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKRSDKRDDS
                     DLVRSELEKPRGRKKTPVTEREEKLGMDDLTKLVQEQRPKGSQRSRKRGHTASESDEQ
                     QWPEEKRLKEDILENEDEQNSPPKKGKRGRPPKPLGGGTPKEEPTMKTSKKGSKKKSG
                     PPAPEEEEEEERQSGNTEQKSKSKQHRVSRRAQQRAESSESSAIESTQSTPQKGRGRP
                     SKTPSPSQPKKNV"
     exon            174..377
                     /gene="AS3"
                     /number=3
     exon            378..464
                     /gene="AS3"
                     /number=4
     exon            465..562
                     /gene="AS3"
                     /number=5
     exon            563..689
                     /gene="AS3"
                     /number=6
     exon            690..770
                     /gene="AS3"
                     /number=7
     exon            771..911
                     /gene="AS3"
                     /number=8
     exon            912..1027
                     /gene="AS3"
                     /number=9
     exon            1028..1122
                     /gene="AS3"
                     /number=10
     exon            1123..1268
                     /gene="AS3"
                     /number=11
     exon            1269..1420
                     /gene="AS3"
                     /number=12
     exon            1421..1534
                     /gene="AS3"
                     /number=13
     exon            1535..1616
                     /gene="AS3"
                     /number=14
     exon            1617..1665
                     /gene="AS3"
                     /number=15
     exon            1666..1805
                     /gene="AS3"
                     /number=16
     exon            1806..1921
                     /gene="AS3"
                     /number=17
     exon            1922..2027
                     /gene="AS3"
                     /number=18
     exon            2028..2188
                     /gene="AS3"
                     /number=19
     exon            2189..2312
                     /gene="AS3"
                     /number=20
     exon            2313..2471
                     /gene="AS3"
                     /number=21
     exon            2472..2540
                     /gene="AS3"
                     /number=22
     exon            2541..2677
                     /gene="AS3"
                     /number=23
     exon            2678..2801
                     /gene="AS3"
                     /number=24
     exon            2802..3006
                     /gene="AS3"
                     /number=25
     exon            3007..3121
                     /gene="AS3"
                     /number=26
```

```
     exon            3122..3254
                     /gene="AS3"
                     /number=27
     exon            3255..3374
                     /gene="AS3"
                     /number=28
     exon            3375..3437
                     /gene="AS3"
                     /number=29
     exon            3438..3583
                     /gene="AS3"
                     /number=30
     exon            3584..3689
                     /gene="AS3"
                     /number=31
     exon            3690..4129
                     /gene="AS3"
                     /number=32
     exon            4130..4354
                     /gene="AS3"
                     /number=33
     exon            4355..4810
                     /gene="AS3"
                     /number=34
BASE COUNT     1755 a    944 c   1074 g   1480 t
ORIGIN

Alignment Scores:
Pred. No.:              0              Length:         5253
Score:                  7101.00        Matches:        1377
Percent Similarity:     99.28%         Conservative:   4
Best Local Similarity:  98.99%         Mismatches:     10
Query Match:            98.72%         Indels:         0
DB:                     9              Gaps:           0

US-09-512-581B-2 (1-1391) x HSU95825 (1-5253)

Qy          1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProProGlyValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 125

Qy         21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValValLys 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 185

Qy         41 ThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 245

Qy         61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
              ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db        246 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGATAAAGATGTTCGCTTACTGGTA 305

Qy         81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 365

Qy        101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 425

Qy        121 LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 AAGAGCCCACAGTTCAATAGGTATTTTTATTTGCTTGAGAACATTGCTTGGGTCAAGTCA 485

Qy        141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 545

Qy        161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 605

Qy        181 MetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 665

Qy        201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 725

Qy        221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
              ||||||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db        726 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTTTTAATCAGGTT 785

Qy        241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeu 260
              |||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db        786 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTCTGACTTAATTTTG 845

Qy        261 GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 905

Qy        281 LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        906 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCGAAAATGTTT 965

Qy        301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGly 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        966 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 1025

Qy        321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1026 AGGTTTAATGATATCCATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1085

Qy        341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1086 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1145

Qy        361 AspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1146 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1205

Qy        381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
              |||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db       1206 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGGGAGAGAGAACATTAGACAAA 1265

Qy        401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAla 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1266 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1325

Qy        421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1326 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1385

Qy        441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1386 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1445

Qy        461 GlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1446 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1505

Qy        481 TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1506 TATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATGTGGAAATGTCAAAAT 1565

Qy        501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1566 CTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAACCCAAAACAGATGCC 1625

Qy        521 SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly 540
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db       1626 AGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAGTTTACCTGATCCTGGT 1685

Qy        541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArg 560
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1686 AAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGATGATGAGAAAATAAGA 1745

Qy         561 LysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysVal 580
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1746 AAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAGGCTGAAGGTTGTGTG 1805

Qy         581 ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1806 CGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAATCCTTTCCTGGAAATG 1865

Qy         601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla 620
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1866 ATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACCGAATCTATCAGTGCT 1925

Qy         621 LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyVal 640
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1926 CTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGATGAAGATGAGGGTGTT 1985

Qy         641 ProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHis 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1986 CCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTACTCTCATTTACACAT 2045

Qy         661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2046 CCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCCTGTCTGAAAATGGAT 2105

Qy         681 AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2106 GATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2165

Qy         701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLys 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2166 GAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACATCACAAATCTAAAAAA 2225

Qy         721 GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys 740
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2226 GGACCCCCCCGTCAAGCCAAATATGCCATTCATTGTATCCATGCGATATTTTCTAGTAAA 2285

Qy         741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
               |||    |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2286 GAGGCCCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCTAGATCCAAGCAACCTG 2345

Qy         761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2346 GAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCTCCTTGCACCTGATCAA 2405

Qy         781 PheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
               |||||||||||    ||||||   ||||||||||||||||||||||||||||||||||||
Db        2406 TTTGCTGCTCCTTTGAAATCTTTGGTAGCTACTTTCATTGTGAAAGATCTTCTCATGAAT 2465

Qy         801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerPro 820
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2466 GATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGATGAAGAAGTATCTCCT 2525

Qy         821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLys 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2526 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAA 2585

Qy         841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2586 AATAATCACAGTAAATCAGGAACTTCTACCTTAAGATTGCTAACAACAATATTGCATAGT 2645

Qy         861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2646 GATGGAGACTTGACAGAACAGGGGAAAATTAGTAAACCAGATATGTCACGTCTGAGACTT 2705

Qy         881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2706 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAAGAACCCTGTTACCATGAAATCATCACA 2765

Qy         901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2766 TTAGAACAATATCAGCTATGTGCATTAGCTATCAACGATGAATGCTATCAAGTAAGACAA 2825

Qy         921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet 940
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2826 GTGTTTGCCCAGAAACTTCACAAAGGCCTTTCCCGTTTACGGCTTCCACTTGAGTATATG 2885

Qy         941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGln 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2886 GCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGAAGAGCTCATGCTAGGCAA 2945

Qy         961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSer 980
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2946 TGTTTGGTGAAAAATATAAATGTAAGGCGGGAGTATCTGAAGCAGCATGCAGCTGTTAGT 3005

Qy         981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3006 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3065

Qy        1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3066 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTTAAAGAATGT 3125

Qy        1021 LeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle 1040
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3126 CTTTGGTTTGTTCTGGAAATATTAATGGCTAAAAATGAAAATAACAGTCACGCTTTTATC 3185

Qy        1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3186 AGAAAGATGGTAGAAAATATTAAGCAAACAAAAGATGCCCAAGGACCAGATGATGCAAAA 3245

Qy        1061 MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3246 ATGAATGAAAAACTGTACACTGTGTGTGATGTTGCCATGAATATCATCATGTCAAAGAGT 3305

Qy        1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln 1100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3306 ACTACATACAGTTTGGAATCTCCTAAAGACCCGGTACTACCAGCTCGTTTCTTCACTCAA 3365

Qy        1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe 1120
               ||||||||||||||||||||||||||||||||||||||||     |||||||||||||||
Db        3366 CCTGACAAGAATTTCAGTAACACCAAAAATTATCTGCCTCCTGGAATGAAATCATTTTTC 3425

Qy        1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3426 ACTCCTGGAAAACCTAAAACAACCAATGTTCTAGGAGCTGTTAACAAGCCACTTTCATCA 3485

Qy        1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
               ||||||||||||||||||||||||||||||||||||||||||||    ||||||||||||
Db        3486 GCAGGCAAGCAATCTCAGACCAAATCATCACGAATGGAAACTGTAGGCAATGCAAGCAGC 3545

Qy        1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3546 AGCTCAAATCCAAGCTCTCCTGGAAGAATAAAGGGGAGGCTTGATAGTTCTGAAATGGAT 3605

Qy        1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
               |||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        3606 CACAGTGAAAATGAAGATTACACAATGTCTTCACCTTTGCCGGGGAAAAGAAGTGACAAG 3665

Qy        1201 ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThr 1220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3666 AGAGACGACTCTGATCTTGTAAGGTCTGAATTGGAGAAGCCTAGAGGCAGAAAAAAAACG 3725

Qy        1221 ProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlu 1240
               ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db        3726 CCCGTCACAGAACGGGAGGAGAAATTAGGTATGGATGACTTGACTAAGTTGGTACAGGAA 3785

Qy        1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp 1260
               |||::::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3786 CAGAGACCTAAAGGCAGTCAGCGAAGTCGGAAAAGAGGCCATACGGCTTCAGAATCTGAT 3845

Qy        1261 GluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGlu 1280
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3846 GAACAGCAGTGGCCTGAGGAAAAGAGGCTCAAAGAAGATATATTAGAAAATGAAGATGAA 3905
```

```
Qy       1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3906 CAGAATAGTCCGCCAAAAAAGGGTAAAAGAGGCCGACCACCAAAACCTCTTGGTGGAGGT 3965

Qy       1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLysSerGly 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3966 ACACCAAAAGAAGAGCCAACAATGAAAACTTCTAAAAAAGGAAGCAAAAAAAAATCTGGA 4025

Qy       1321 ProProAlaProGluGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys 1340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4026 CCTCCAGCACCAGAGGAGGAGGAAGAAGAAGAAAGACAAAGTGGAAATACGGAACAGAAG 4085

Qy       1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||    |||
Db       4086 TCCAAAAGCAAACAGCACCGAGTGTCAAGGAGAGCACAGCAGAGAGCAGAATCTTCTGAA 4145

Qy       1361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4146 TCTAGTGCAATTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4205

Qy       1381 ThrProSerProSerGlnProLysLysAsnVal 1391
              |||||||||||||||||||||||||||||||||
Db       4206 ACGCCATCACCATCACAACCAAAAAAAAATGTG 4238

RESULT 4
AY102267
LOCUS       AY102267                5317 bp    mRNA    linear   ROD 01-JUN-2003
DEFINITION  Mus musculus androgen-induced prostate proliferative shutoff
            associated protein AS3 mRNA, complete cds.
ACCESSION   AY102267
VERSION     AY102267.1  GI:31321922
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 5317)
  AUTHORS   Geck,P., Maffini,M., Sonnenschein,C. and Soto,A.M.
  TITLE     The AS3 proliferative arrest gene has an ancient eukaryotic
            heritage and shows highly conserved functional domains in mice
  JOURNAL   Proc. Annu. Meet. Am. Assoc. Cancer Res. 43, 987 (2002)
REFERENCE   2  (bases 1 to 5317)
  AUTHORS   Geck,P., Maffini,M., Liang,S.L., Sonnenschein,C. and Soto,A.M.
  TITLE     Extreme domain conservation and expression of the mouse AS3
            proliferative arrest protein
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5317)
  AUTHORS   Geck,P., Maffini,M., Liang,S.L., Sonnenschein,C. and Soto,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2002) Anatomy and Cell Biology, Tufts University
            School of Medicine, 136 Harrison Avenue, Boston, MA 02111, USA
FEATURES             Location/Qualifiers
     source          1..5317
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /map="84cM; G2; between BRCA2 and KL"
                     /sex="male"
                     /tissue_type="brain"
     CDS             147..4487
                     /codon_start=1
                     /product="androgen-induced prostate proliferative shutoff
                     associated protein AS3"
                     /protein_id="AAM52216.1"
                     /db_xref="GI:31321923"
                     /translation="MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMD
                     MDQDSEEEKELYLNLALHLASDFFLKHPDKDVRLLVACCLADIFRIYAPEAPYTSPDK
                     LKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRT
                     LFSVINNGHNQKVHMHMVDLMSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLA
                     KALLKRTAQAIEPYITNFFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQ
                     LEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECV
                     KFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNF
                     VRERTLDKRWRVRKEAMMGLAQIYKKYSLQSAAGKDAAKQISWVKDKLLHIYYQNSID
                     DRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKALNEMRKCQNLLRHQVKD
                     LLDLIKQPKTDASVKAIFSKVMVITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEAL
                     VSPTCSCKQAEGCVREITKKLGNPKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQ
                     VNKSIDGTADDEDEGVPTDQAIRAGLELLKVLSFTHPISFHSAETFESLLACLKMDDE
                     KVAEAALQIFKNTGSKIEEDFPHIRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSK
                     ETQFAQIFEPLHKSLDPSNLEHLITPLVTIGHIALLAPDQFAAPLKSLVATFIVKDLL
                     MNDRLPGKKTTKLWVPDEEVSPETMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTT
                     ILHSDGDLTEQGKISKPDMSRLRLAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDE
                     CYQVRQVFAQKLHKGLSRLRLPLEYMAICALCAKDPVKERRAHARQCLVKNTTVRREY
                     LKQHASVSEKLLSLLPEYVVPYTIHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMA
                     KNENNSHAFIRKMVENIKQTKDAQGPDDTKMNEKLYTVCDVAMNIIMSKSTTYSLESP
                     KDPVLPARFFTQPDKNFSNTKNYLPPEMKSFFTPGKPKTANVLGAVNKPLSSAGKQSQ
                     TKSSRMETVSNASSSSNPSSPGRIKGRLDSSEMDHSENEDYTMSSPLPGKKSDKREDP
                     DLSELEKPRSRKKAPVTDPEEKLGMDDLTKLVQEQKPKGSQRGRKRGRTASDSDEQQW
                     PEEKRHKEELLENEDEQNSPPKKGKRGRPPKPLGGGTSKEEPTMKTSKKGNKKKLVPP
                     VVDDEEEERQIGNTEHKSKSKQHRTSKRAQQRAESPETSAVESTQSTPQKGRGRPSK
                     APSPSQPPKKIRVGRSKQVATKENDSSEEMDVLQASSPVSDDTTQEGAEEEDISVGNV
                     RRRSSKRERR"
BASE COUNT     1676 a   1100 c   1199 g   1342 t
ORIGIN

Alignment Scores:
Pred. No.:              0            Length:        5317
Score:                  6942.00      Matches:       1344
Percent Similarity:     97.84%       Conservative:  17
Best Local Similarity:  96.62%       Mismatches:    28
Query Match:            96.51%       Indels:        2
DB:                     10           Gaps:          1

US-09-512-581B-2 (1-1391) x AY102267 (1-5317)

Qy          1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProProGlyValLys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 ATGGCTCATTCAAAGACAAGGACCAACGATGGGAAAATTACTTACCCTCCTGGAGTCAAG 206

Qy         21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValValLys 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 GAAATCTCAGATAAAATCTCTAAAGAGGAGATGGTGAGGCGGTTAAAGATGGTTGTAAAA 266

Qy         41 ThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 ACTTTCATGGACATGGACCAGGACTCTGAAGAGGAAAAGGAACTTTATCTAAACCTAGCT 326

Qy         61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
              ||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db        327 TTACATCTTGCTTCTGACTTCTTCCTCAAGCATCCTGATAAAGATGTTCGTTTACTGGTG 386

Qy         81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 GCTTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAGGCTCCTTACACGTCTCCC 446

Qy        101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 GATAAACTAAAGGATATATTTATGTTTATAACAAGGCAACTAAAGGGACTAGAAGATACA 506

Qy        121 LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 AAGAGCCCTCAGTTTAATAGGTATTTTTATTTACTTGAGAACATTGCATGGGTTAAATCA 566

Qy        141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        567 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATCTTTACTCAACTTTACAGAACA 626

Qy        161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        627 TTATTCTCAGTTATAAACAATGGCCACAATCAGAAAGTTCATATGCACATGGTGGACCTC 686

Qy        181 MetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        687 ATGAGTTCTATCATTTGTGAAGGTGATACCGTATCTCAGGAGCTCTTAGATACAGTTTTA 746
```

NPD

```
Qy        201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        747 GTAAATCTGGTACCTGCCCATAAGAACTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 806

Qy        221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
              ||||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        807 TTGCTGAAGAGGACTGCTCAAGCTATTGAACCATATATTACCAATTTTTTTAACCAGGTT 866

Qy        241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        867 CTGATGCTTGGGAAAACATCTATCAGTGATTTGTCTGAGCATGTCTTTGATTTGATCTTG 926

Qy        261 GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        927 GAACTCTACAACATTGACAGCCATTTGCTGCTTTCTGTTCTGCCCCAGCTTGAGTTTAAA 986

Qy        281 LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        987 TTAAAGAGCAATGATAATGAAGAACGCCTCCAGGTTGTTAAACTACTGGCGAAAATGTTT 1046

Qy        301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGly 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1047 GGGGCAAAGGATTCAGAATTGGCTTCTCAGAACAAACCCCTCTGGCAGTGCTACCTGGGC 1106

Qy        321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1107 AGGTTTAATGATATCCATGTACCAATACGCCTGGAATGTGTGAAATTTGCCAGCCACTGC 1166

Qy        341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1167 CTCATGAACCATCCTGATTTAGCAAAGGATTTAACAGAGTATCTTAAAGTGAGGTCACAT 1226

Qy        361 AspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1227 GATCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCTATAGTAACAGCTGCTAAAAAG 1286

Qy        381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1287 GATATTCTTCTTGTCAATGATCACTTACTCAATTTTGTGAGAGAGCGAACATTGGACAAA 1346

Qy        401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAla 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db       1347 CGGTGGAGAGTACGTAAAGAAGCCATGATGGGACTGGCCCAGATTTATAAGAAATACTCT 1406

Qy        421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu 440
              |||||||||||||||||||||||||||||||||||||||||:::|||:::||||||||||
Db       1407 TTACAGTCAGCAGCTGGCAAAGATGCTGCAAAACAGATATCCTGGGTCAAGGACAAGCTG 1466

Qy        441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1467 CTGCATATATACTATCAGAACAGCATCGACGACCGGCTACTTGTTGAGCGGATCTTTGCT 1526

Qy        461 GlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1527 CAGTACATGGTCCCTCACAATTTGGAAACGACGGAACGGATGAAGTGCTTGTATTACTTG 1586

Qy        481 TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsn 500
              |||||||||||||||||||||||||||||||||||||||||||||    |||||||||||
Db       1587 TACGCTACACTGGACTTGAATGCTGTGAAAGCATTGAATGAAATGCGGAAATGTCAAAAT 1646

Qy        501 LeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnProLysThrAspAla 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1647 CTACTCCGACATCAAGTAAAGGATTTACTTGATTTAATTAAACAACCAAAAACAGATGCC 1706

Qy        521 SerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeuProAspProGly 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1707 AGTGTCAAGGCCATATTTTCTAAAGTGATGGTTATCACAAGAAATTTGCCAGATCCTGGT 1766

Qy        541 LysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArg 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1767 AAGGCCCAAGACTTCATGAAGAAGTTCACACAAGTGTTGGAAGATGATGAGAAAATACGG 1826

Qy        561 LysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysVal 580
              |||||||||||    |||||||||||||||||||||||||||||||||||||||||||||
Db       1827 AAACAGCTGGAAGCCCTTGTTAGTCCCACATGTTCCTGCAAGCAGGCTGAAGGCTGTGTG 1886

Qy        581 ArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMet 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1887 CGTGAAATAACTAAGAAGTTGGGCAACCCTAAGCAGCCTACAAACCCTTTCTTGGAAATG 1946

Qy        601 IleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAla 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1947 ATCAAGTTTCTCTTGGAGAGGATAGCTCCTGTACACATAGACACTGAATCCATCAGTGCT 2006

Qy        621 LeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyVal 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2007 CTTATCAAGCAAGTGAATAAATCAATAGATGGAACAGCAGATGATGAGGATGAGGGTGTT 2066

Qy        641 ProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHis 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2067 CCAACTGATCAAGCTATCAGGGCAGGCCTTGAGCTGCTGAAGGTGCTGTCTTTTACGCAC 2126

Qy        661 ProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAsp 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2127 CCTATCTCCTTTCACTCTGCTGAGACGTTTGAGTCCCTTCTGGCTTGTCTGAAGATGGAT 2186

Qy        681 AspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGlu 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2187 GATGAGAAGGTGGCAGAAGCTGCACTACAAATTTTCAAAAACACAGGAAGCAAAATTGAA 2246

Qy        701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLys 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2247 GAGGATTTCCCTCACATCAGATCAGCCTTGCTTCCTGTACTACATCACAAATCCAAAAAA 2306

Qy        721 GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLys 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2307 GGACCACCTCGGCAGGCCAAATACGCCATCCATTGTATTCATGCCATATTTTCTAGTAAA 2366

Qy        741 GluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2367 GAAACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAAAGCCTAGATCCAAGCAACCTG 2426

Qy        761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2427 GAGCATCTTATAACCCCTCTGGTCACTATTGGTCATATTGCTCTTCTCGCACCTGATCAG 2486

Qy        781 PheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
              ||||||||||||    ||||||    ||||||||||||||||||||||||||||||||||
Db       2487 TTCGCTGCTCCTCTGAAGTCTTTGGTGGCAACTTTCATTGTGAAGGACCTCCTCATGAAT 2546

Qy        801 AspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGluGluValSerPro 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2547 GACCGGCTTCCAGGAAAAAAGACAACTAAGCTTTGGGTTCCAGATGAGGAAGTCTCACCT 2606

Qy        821 GluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLys 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2607 GAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGGCTACTTGGAATGAAG 2666

Qy        841 AsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSer 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2667 AATAATCACAGTAAGTCAGGAACTTCCACCCTCAGATTACTAACAACGATACTGCATAGT 2726

Qy        861 AspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSerArgLeuArgLeu 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2727 GATGGGGATTTGACAGAACAAGGGAAAATTAGTAAACCAGATATGTCACGCCTGAGACTT 2786

Qy        881 AlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHisGluIleIleThr 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2787 GCTGCTGGGAGTGCTATTGTGAAGCTGGCACAGGAGCCCTGTTACCACGAGATCATTACA 2846

Qy        901 LeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGln 920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2847 CTGGAGCAGTACCAGCTGTGTGCATTAGCCATCAATGATGAGTGTTATCAAGTCAGGCAG 2906

Qy        921 ValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMet 940
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2907 GTGTTCGCCCAGAAACTTCACAAAGGCCTTTCCCGCTTACGGCTTCCCCTTGAGTACATG 2966

Qy         941 AlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAlaHisAlaArgGln 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2967 GCCATCTGTGCTCTTTGTGCCAAAGACCCTGTGAAGGAGAGGCGAGCCCATGCTAGACAG 3026

Qy         961 CysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSer 980
               |||||||||||||||      ||||||||||||||||||||||||||||||:::||||||
Db        3027 TGCTTGGTGAAAAACACCACTGTGAGGAGGGAGTACCTGAAGCAGCATGCAAGTGTTAGT 3086

Qy         981 GluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAla 1000
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3087 GAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACAATTCACCTTTTGGCA 3146

Qy        1001 HisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCys 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3147 CATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAAGATGTGAAAGAATGC 3206

Qy        1021 LeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIle 1040
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3207 CTTTGGTTTGTTCTGGAGATATTGATGGCTAAAAATGAAAACAACAGCCATGCATTTATC 3266

Qy        1041 ArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLys 1060
               ||||||||||||||||||||||||||||||||||||||||||||||||||||    |||
Db        3267 AGAAAAATGGTAGAAAATATTAAACAGACAAAAGATGCTCAAGGACCAGATGATACAAAA 3326

Qy        1061 MetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysSer 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3327 ATGAATGAAAAATTGTACACCGTGTGTGATGTTGCCATGAACATCATCATGTCCAAGAGC 3386

Qy        1081 ThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrGln 1100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3387 ACCACGTACAGCCTGGAGTCTCCTAAGGACCCCGTGCTGCCAGCTCGGTTTTTCACCCAG 3446

Qy        1101 ProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMetLysSerPhePhe 1120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3447 CCTGACAAGAATTTTAGTAACACCAAAAATTACCTGCCTCCAGAAATGAAATCATTTTTC 3506

Qy        1121 ThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSer 1140
               ||||||||||||||||||    ||||||||||||||||||||||||||||||||||||||
Db        3507 ACTCCTGGAAAACCTAAAACAGCCAATGTTCTCGGAGCAGTTAATAAGCCACTTTCATCA 3566

Qy        1141 AlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSerAsnAlaSerSer 1160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3567 GCAGGCAAACAGTCTCAGACCAAATCATCAAGAATGGAAACTGTGAGCAACGCAAGCAGC 3626

Qy        1161 SerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAsp 1180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3627 AGCTCCAACCCAAGCTCTCCTGGAAGGATCAAGGGGAGGCTTGATAGCTCTGAAATGGAT 3686

Qy        1181 HisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3687 CACAGTGAAAATGAAGATTATACAATGTCTTCACCTTTGCCAGGAAAAAAAAGTGACAAG 3746

Qy        1201 ArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThr 1220
               |||:::|||  ||||||      |||||||||||||||||||||    ||||||||
Db        3747 AGAGAAGACCCTGATCTT------TCTGAGTTGGAGAAGCCTAGAAGTCGGAAAAAAGCA 3800

Qy        1221 ProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGlu 1240
               |||||||||:::   |||||||||||||||||||||||||||||||||||||||||||||
Db        3801 CCTGTCACAGACCCTGAAGAGAAATTAGGTATGGATGACCTAACTAAGTTGGTACAGGAA 3860

Qy        1241 GlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAsp 1260
               ||||||||||||||||||||||||     ||||||||||||    ||||||||:::||||||
Db        3861 CAGAAACCTAAAGGCAGTCAGCGAGGCCGGAAAAGAGGCCGTACAGCTTCAGACTCAGAC 3920

Qy        1261 GluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGlu 1280
               |||||||||||||||||||||||||||    ||||||:::::::||||||||||||||||||
Db        3921 GAGCAGCAGTGGCCTGAGGAGAAGAGGCACAAAGAGGAGCTCCTGGAAAATGAGGATGAG 3980

Qy        1281 GlnAsnSerProProLysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGly 1300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3981 CAGAACAGCCCACCAAAAAAGGGCAAAAGAGGCAGGCCACCAAAACCTCTTGGTGGGGGG 4040

Qy        1301 ThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLysSerGly 1320
               |||   ||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        4041 ACATCGAAGGAAGAGCCCACAATGAAAACATCCAAGAAAGGAAACAAGAAAAAACTTGTA 4100

Qy        1321 ProProAlaProGluGluGluGluGluGluGluArgGlnSerGlyAsnThrGluGlnLys 1340
               ||||||       :::::::::|||||||||||||||||    ||||||||||||    |||
Db        4101 CCTCCTGTAGTAGACGACGATGAAGAAGAAGAAAGACAAATTGGAAACACAGAACATAAG 4160

Qy        1341 SerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArgAlaGluSerProGlu 1360
               ||||||||||||||||||||    ||||:::||||||||||||||||||||||||||||||
Db        4161 TCAAAAAGCAAACAGCACCGCACATCAAAGAGAGCACAACAGAGAGCAGAATCTCCTGAA 4220

Qy        1361 SerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArgGlyArgProSerLys 1380
               :::||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        4221 ACAAGTGCAGTTGAATCCACACAGTCCACACCACAGAAAGGACGAGGAAGACCATCAAAA 4280

Qy        1381 ThrProSerProSerGlnProLysLysAsnVal 1391
                  ||||||||||||||||   |||    :::
Db        4281 GCACCATCACCATCACAACCCCCCAAAAAAATC 4313
```

NM

```
RESULT 5
AK122414
LOCUS       AK122414                5944 bp    mRNA    linear   ROD 15-MAR-2003
DEFINITION  Mus musculus mRNA for mKIAA0979 protein.
ACCESSION   AK122414
VERSION     AK122414.1  GI:28972559
KEYWORDS    FLI_CDNA.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Aizawa,H., Yuasa,S.,
            Nakajima,D., Nagase,T., Ohara,O. and Koga,H.
  TITLE     Prediction of the coding sequences of mouse homologues of KIAA
            gene: II. The complete nucleotide sequences of 400 mouse
            KIAA-homologous cDNAs identified by screening of terminal sequences
            of cDNA clones randomly sampled from size-fractionated libraries
  JOURNAL   DNA Res. 10, 35-48 (2003)
REFERENCE   2  (bases 1 to 5944)
  AUTHORS   Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research
            Institute, Laboratory for Genome Informatics; 2-6-7
            Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
            (E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
COMMENT     The CREATE program supported by Japan science and technology
            corporation; cDNA full insert sequencing: Kazusa DNA Research
            Institute; cDNA library construction, clone selection and 5'- &
            3'-end one pass sequencing.
FEATURES             Location/Qualifiers
     source          1..5944
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="mbx00002"
                     /tissue_type="brain"
                     /dev_stage="adult"
                     /note="vector:modified pBC SK+"
     gene            1..5944
                     /gene="mKIAA0979"
     CDS             <1..3577
                     /gene="mKIAA0979"
                     /note="CDS is predicted by in silico analysis. Start codon
                     is not identified."
                     /codon_start=2
                     /evidence=not_experimental
                     /product="mKIAA0979 protein"
                     /protein_id="BAC65696.1"
                     /db_xref="GI:28972560"
                     /translation="FDLILELYNIDSHLLLSVLPQLEFKLKSNDNEERLQVVKLLAKM
```

```
                          FGAKDSELASQNKPLWQCYLGRFNDIHVPIRLECVKFASHCLMNHPDLAKDLTEYLKV
                          RSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNFVRERTLDKRWRVRKEAMMGLAQI
                          YKKYSLQSAAGKDAAKQISWVKDKLLHIYYQNSIDDRLLVERIFAQYMVPHNLETTER
                          MKCLYYLYATLDLNAVKALNEMWKCQNLLRHQVKDLLDLIKQPKTDASVKAIFSKVMV
                          ITRNLPDPGKAQDFMKKFTQVLEDDEKIRKQLEALVSPTCSCKQAEGCVREITKKLGN
                          PKQPTNPFLEMIKFLLERIAPVHIDTESISALIKQVNKSIDGTADDEDEGVPTDQAIR
                          AGLELLKVLSFTHPISFHSAETFESLLACLKMDDEKVAEAALQIFKNTGSKIEEDFPH
                          IRSALLPVLHHKSKKGPPRQAKYAIHCIHAIFSSKETQFAQIFEPLHKSLDPSNLEHL
                          ITPLVTIGHIALLAPDQFAAPLKSLVATFIVKDLLMNDRLPGKKTTKLWVPDEEVSPE
                          TMVKIQAIKMMVRWLLGMKNNHSKSGTSTLRLLTTILHSDGDLTEQGKISKPDMSRLR
                          LAAGSAIVKLAQEPCYHEIITLEQYQLCALAINDECYQVRQVFAQKLHKGLSRLRLPL
                          EYMAICALCAKDPVKERRAHARQCLVKNITVRREYLKQHAAVSEKLLSLLPEYVVPYT
                          IHLLAHDPDYVKVQDIEQLKDVKECLWFVLEILMAKNENNSHAFIRKMVENIKQTKDA
                          QGPDDTKMNEKLYTVCDVAMNIIMSKSTTYSLESPKDPVLPARFFTQPDKNFSNTKNY
                          LPPEMKSFFTPGKPKTANVLGAVNKPLSSAGKQSQTKSSRMETVSNASSSSNPSSPGR
                          IKGRLDSSEMDHSENEDYTMSSPLPGKKSDKREDPDLSELEKPRSRKKAPVTDPEEKL
                          GMDDLTKLVQEQKPKGSQRGRKRGRTASDSDEQQWPEEKRHKEELLENEDEQNSPPKK
                          GKRGRPPKPLGGGTSKEEPTMKTSKKGNKKKLVPPVVDDDEEEERQIGNTEHKSKSKQ
                          HRTSKRAQQRAESPETSAVESTQSTPQKGRGRPSKAPSPSQPPKKIRVGRSKQVATKE
                          NDSSEEMDVLQASSPVSDDTTQEGAEEEDISVGNVRRRSSKRERR"
BASE COUNT     1892 a   1151 c   1251 g   1650 t
ORIGIN

Alignment Scores:
Pred. No.:              1.84e-290       Length:         5944
Score:                  5659.00         Matches:        1094
Percent Similarity:     97.71%          Conservative:   16
Best Local Similarity:  96.30%          Mismatches:     24
Query Match:            78.67%          Indels:         2
DB:                     10              Gaps:           1

US-09-512-581B-2 (1-1391) x AK122414 (1-5944)

Qy        256 PheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuPro 275
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 TTTGATTTGATCTTGGAACTCTACAACATTGACAGCCATTTGCTGCTTTCTGTTCTGCCC 61

Qy        276 GlnLeuGluPheLysLeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeu 295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CAGCTTGAGTTTAAATTAAAGAGCAATGATAATGAAGAACGCCTCCAGGTTGTTAAACTA 121

Qy        296 LeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrp 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 CTGGCGAAAATGTTTGGGGCAAAGGATTCAGAATTGGCTTCTCAGAACAAACCCCTCTGG 181

Qy        316 GlnCysTyrLeuGlyArgPheAsnAspIleHisValProIleArgLeuGluCysValLys 335
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 CAGTGCTACCTGGGCAGGTTTAATGATATCCATGTACCAATACGCCTGGAATGTGTGAAA 241

Qy        336 PheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeu 355
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TTTGCCAGCCACTGCCTCATGAACCATCCTGATTTAGCAAAGGATTTAACAGAGTATCTT 301

Qy        356 LysValArgSerHisAspProGluGluAlaIleArgHisAspValIleValSerIleVal 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 AAAGTGAGGTCACATGATCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCTATAGTA 361

Qy        376 ThrAlaAlaLysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGlu 395
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 ACAGCTGCTAAAAAGGATATTCTTCTTGTCAATGATCACTTACTCAATTTTGTGAGAGAG 421

Qy        396 ArgThrLeuAspLysArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIle 415
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 CGAACATTGGACAAACGGTGGAGAGTACGTAAAGAAGCCATGATGGGACTGGCCCAGATT 481

Qy        416 TyrLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrp 435
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||:::|||
Db        482 TATAAGAAATACTCTTTACAGTCAGCAGCTGGCAAAGATGCTGCAAAACAGATATCCTGG 541

Qy        436 IleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuVal 455
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 GTCAAGGACAAGCTGCTGCATATATACTATCAGAACAGCATCGACGACCGGCTACTTGTT 601

Qy        456 GluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLys 475
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 GAGCGGATCTTTGCTCAGTACATGGTCCCTCACAATTTGGAAACGACGGAACGGATGAAG 661

Qy        476 CysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMet 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        662 TGCTTGTATTACTTGTACGCTACACTGGACTTGAATGCTGTGAAAGCATTGAATGAAATG 721

Qy        496 TrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGln 515
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 TGGAAATGTCAAAATCTACTCCGACATCAAGTAAAGGATTTACTTGATTTAATTAAACAA 781

Qy        516 ProLysThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsn 535
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 CCAAAAACAGATGCCAGTGTCAAGGCCATATTTTCTAAAGTGATGGTTATCACAAGAAAT 841

Qy        536 LeuProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAsp 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 TTGCCAGATCCTGGTAAGGCCCAAGACTTCATGAAGAAGTTCACACAAGTGTTGGAAGAT 901

Qy        556 AspGluLysIleArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGln 575
              ||||||||||||||||||||||||||      ||||||||||||||||||||||||||||
Db        902 GATGAGAAAATACGGAAACAGCTGGAAGCCCTTGTTAGTCCCACATGTTCCTGCAAGCAG 961

Qy        576 AlaGluGlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsn 595
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 GCTGAAGGCTGTGTGCGTGAAATAACTAAGAAGTTGGGCAACCCTAAGCAGCCTACAAAC 1021

Qy        596 ProPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAspThr 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1022 CCTTTCTTGGAAATGATCAAGTTTCTCTTGGAGAGGATAGCTCCTGTACACATAGACACT 1081

Qy        616 GluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAsp 635
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1082 GAATCCATCAGTGCTCTTATCAAGCAAGTGAATAAATCAATAGATGGAACAGCAGATGAT 1141

Qy        636 GluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysVal 655
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1142 GAGGATGAGGGTGTTCCAACTGATCAAGCTATCAGGGCAGGCCTTGAGCTGCTGAAGGTG 1201

Qy        656 LeuSerPheThrHisProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAla 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1202 CTGTCTTTTACGCACCCTATCTCCTTTCACTCTGCTGAGACGTTTGAGTCCCTTCTGGCT 1261

Qy        676 CysLeuLysMetAspAspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThr 695
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1262 TGTCTGAAGATGGATGATGAGAAGGTGGCAGAAGCTGCACTACAAATTTTCAAAAACACA 1321

Qy        696 GlySerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHis 715
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1322 GGAAGCAAAATTGAAGAGGATTTCCCTCACATCAGATCAGCCTTGCTTCCTGTACTACAT 1381

Qy        716 HisLysSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAla 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1382 CACAAATCCAAAAAAGGACCACCTCGGCAGGCCAAATACGCCATCCATTGTATTCATGCC 1441

Qy        736 IlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeu 755
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1442 ATATTTTCTAGTAAAGAAACCCAGTTTGCACAGATATTTGAGCCTCTGCATAAAAGCCTA 1501

Qy        756 AspProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeu 775
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1502 GATCCAAGCAACCTGGAGCATCTTATAACCCCTCTGGTCACTATTGGTCATATTGCTCTT 1561

Qy        776 LeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLys 795
              ||||||||||||||||||||||||||     ||||||      ||||||||||||||||||
Db       1562 CTCGCACCTGATCAGTTCGCTGCTCCTCTGAAGTCTTTGGTGGCAACTTTCATTGTGAAG 1621

Qy        796 AspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAsp 815
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1622 GACCTCCTCATGAATGACCGGCTTCCAGGAAAAAAGACAACTAAGCTTTGGGTTCCAGAT 1681

Qy        816 GluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrp 835
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1682 GAGGAAGTCTCACCTGAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATGG 1741

Qy        836 LeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThr 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1742 CTACTTGGAATGAAGAATAATCACAGTAAGTCAGGAACTTCCACCCTCAGATTACTAACA 1801

Qy        856 ThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMet 875
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1802 ACGATACTGCATAGTGATGGGGATTTGACAGAACAAGGGAAAATTAGTAAACCAGATATG 1861

Qy        876 SerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyr 895
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1862 TCACGCCTGAGACTTGCTGCTGGGAGTGCTATTGTGAAGCTGGCACAGGAGCCCTGTTAC 1921

Qy        896 HisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCys 915
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1922 CACGAGATCATTACACTGGAGCAGTACCAGCTGTGTGCATTAGCCATCAATGATGAGTGT 1981

Qy        916 TyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeu 935
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1982 TATCAAGTCAGGCAGGTGTTCGCTCAGAAACTTCACAAAGGCCTTTCCCGCTTACGGCTT 2041

Qy        936 ProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArg 955
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2042 CCCCTTGAGTACATGGCCATCTGTGCTCTTTGTGCCAAAGACCCTGTGAAAGAGAGGCGA 2101

Qy        956 AlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGln 975
              ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db       2102 GCCCATGCTAGACAGTGCTTGGTGAAGAACATCACTGTGAGGAGGGAGTACCTGAAGCAG 2161

Qy        976 HisAlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThr 995
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2162 CATGCAGCTGTTAGTGAAAAATTATTGTCTCTTCTACCAGAGTATGTTGTTCCATATACA 2221

Qy        996 IleHisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeuLys 1015
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2222 ATTCACCTTTTGGCACATGACCCAGATTATGTCAAAGTACAGGATATTGAACAACTTAAA 2281

Qy       1016 AspValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsn 1035
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2282 GATGTGAAAGAATGCCTTTGGTTTGTTCTGGAGATATTGATGGCTAAAAATGAAAACAAC 2341

Qy       1036 SerHisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGly 1055
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2342 AGCCATGCATTTATCAGAAAAATGGTAGAAAATATTAAACAGACAAAAGATGCTCAAGGA 2401

Qy       1056 ProAspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIle 1075
              ||||||||   |||||||||||||||||||||||||||||||||||||||||||||||||
Db       2402 CCAGATGATACAAAAATGAATGAAAAATTGTACACCGTGTGTGATGTTGCCATGAACATC 2461

Qy       1076 IleMetSerLysSerThrThrTyrSerLeuGluSerProLysAspProValLeuProAla 1095
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2462 ATCATGTCCAAGAGCACCACGTACAGCCTGGAGTCTCCTAAGGACCCCGTGCTGCCAGCT 2521

Qy       1096 ArgPhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGlu 1115
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2522 CGGTTTTTCACCCAGCCTGACAAGAATTTTAGTAACACCAAAAATTACCTGCCTCCAGAA 2581

Qy       1116 MetLysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsn 1135
              ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db       2582 ATGAAATCATTTTTCACTCCTGGAAAACCTAAAACAGCCAATGTTCTCGGAGCAGTTAAT 2641

Qy       1136 LysProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrVal 1155
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2642 AAGCCACTTTCATCAGCAGGCAAACAGTCTCAGACCAAATCATCAAGAATGGAAACTGTG 2701

Qy       1156 SerAsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArgLeuAsp 1175
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2702 AGCAACGCAAGCAGCAGCTCCAACCCAAGCTCTCCTGGAAGGATCAAGGGGAGGCTTGAT 2761

Qy       1176 SerSerGluMetAspHisSerGluAsnGluAspTyrThrMetSerSerProLeuProGly 1195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2762 AGCTCTGAAATGGATCACAGTGAAAATGAAGATTATACAATGTCTTCACCTTTGCCAGGA 2821

Qy       1196 LysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArg 1215
              |||||||||||||||||:::|||  ||||||      |||||||||||||||||||
Db       2822 AAAAAAAGTGACAAGAGAGAAGACCCTGATCTT------TCTGAGTTGGAGAAGCCTAGA 2875

Qy       1216 GlyArgLysLysThrProValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThr 1235
                 ||||||||   |||||||||:::   ||||||||||||||||||||||||||||
Db       2876 AGTCGGAAAAAAGCACCTGTCACAGACCCTGAAGAGAAATTAGGTATGGATGACCTAACT 2935

Qy       1236 LysLeuValGlnGluGlnLysProLysGlySerGlnArgSerArgLysArgGlyHisThr 1255
              ||||||||||||||||||||||||||||||||||||||||   |||||||||||   |||
Db       2936 AAGTTGGTACAGGAACAGAAACCTAAAGGCAGTCAGCGAGGCCGGAAAAGAGGCCGTACA 2995

Qy       1256 AlaSerGluSerAspGluGlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeu 1275
              ||||||:::|||||||||||||||||||||||||||||||||||||||||||:::::|||
Db       2996 GCTTCAGACTCAGACGAGCAGCAGTGGCCTGAGGAGAAGAGGCACAAAGAGGAGCTCCTG 3055

Qy       1276 GluAsnGluAspGluGlnAsnSerProProLysLysGlyLysArgGlyArgProProLys 1295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3056 GAAAATGAGGATGAGCAGAACAGCCCACCAAAAAAGGGCAAAAGAGGCAGGCCACCAAAA 3115

Qy       1296 ProLeuGlyGlyGlyThrProLysGluGluProThrMetLysThrSerLysLysGlySer 1315
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||:::
Db       3116 CCTCTTGGTGGGGGGACATCGAAGGAAGAGCCCACAATGAAAACATCCAAGAAAGGAAAC 3175

Qy       1316 LysLysLysSerGlyProProAlaProGluGluGluGluGluGluArgGlnSerGly 1335
              |||||||||      ||||||      :::::::::|||||||||||||||||   |||
Db       3176 AAGAAAAAACTTGTACCTCCTGTAGTAGACGACGATGAAGAAGAAGAAAGACAAATTGGA 3235

Qy       1336 AsnThrGluGlnLysSerLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArg 1355
              |||||||||   ||||||||||||||||||||||||||   |||:::||||||||||||||
Db       3236 AACACAGAACATAAGTCAAAAAGCAAACAGCACCGCACATCAAAGAGAGCACAACAGAGA 3295

Qy       1356 AlaGluSerProGluSerSerAlaIleGluSerThrGlnSerThrProGlnLysGlyArg 1375
              ||||||||||||:::||||||:::||||||||||||||||||||||||||||||||||||
Db       3296 GCAGAATCTCCTGAAACAAGTGCAGTTGAATCCACACAGTCCACACCACAGAAAGGACGA 3355

Qy       1376 GlyArgProSerLysThrProSerProSerGlnProLysLysAsnVal 1391
              ||||||||||||||   ||||||||||||||||||   |||   :::
Db       3356 GGAAGACCATCAAAAGCACCATCACCATCACAACCCCCCAAAAAAATC 3403
```

RESULT 6
AF294791

```
LOCUS       AF294791                6744 bp    mRNA    linear   PRI 24-JUL-2002
DEFINITION  Homo sapiens SCC-112 (SCC-112) mRNA, complete cds.
ACCESSION   AF294791
VERSION     AF294791.1  GI:21951801
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 6744)
  AUTHORS   Kumar,D., Patel,S., Whiteside,T.L. and Kasid,U.
  TITLE     Identification and characterization of SCC-112, a novel cell cycle
            regulated gene in human cancer cells
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6744)
  AUTHORS   Kumar,D. and Kasid,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2000) Radiation, Medicine, 3970 Reservoir Rd, NW,
            Washington, DC 20007, USA
FEATURES             Location/Qualifiers
     source          1..6744
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
     gene            1..6744
                     /gene="SCC-112"
     CDS             232..4125
                     /gene="SCC-112"
```

```
                     /codon_start=1
                     /product="SCC-112"
                     /protein_id="AAM82347.1"
                     /db_xref="GI:21951802"
                     /translation="MIKRLKMVVKTFMDMDQDSEDEKQQYLPLALHLASEFFLRNPNK
                     DVRLLVACCLADIFRIYAPEAPYTSHDKLKDIFLFITRQLKGLEDTKSPQFNRYFYLL
                     ENLAWVKSYNICFELEDCNEIFIQLFRTLFSVINNSHNKKVQMHMLDLMSSIIMEGDG
                     VTQELLGSILINLIPAHKNLNKQSFDLAKVLLKRTVQTIEACIANFFNQVLVLGRSSV
                     SDLSEHVFDLIQELFAIDPHLLLSVMPQLEFKLKSNDGEERLAVVRLLAKLFGSKDSD
                     LATQNRPLWQCFLGRFNDIHVPVRLESVKFASHCLMNHPDLAKDLTEYLKVRSHDPEE
                     AIRHDVIVTIITAAKRDLALVNDQLLGFVRERTLDKRWRVRKEAMMGLAQLYKKYCLH
                     GEAGKEAAEKVSWIKDKLLHIYYQNSIDDKLLVEKIFAQYLVPHNLETEERMKCLYYL
                     YASLDPNAVKALNEMWKCQNMLRSHVRELLDLHKQPTSEANCSAMFGKLMTIAKNLPD
                     PGKAQDFVKKFNQVLGDDEKLRSQLELLISPTCSCKQADICVREIARKLANPKQPTNP
                     FLEMVKFLLERIAPVHIDSEAISALVKLMNKSIEGTADDEEEGVSPDTAIRSGLELLK
                     VLSFTHPTSFHSAETYESLLQCLRMEDDKVAEAAIQIFRNTGHKIETDLPQIRSTLIP
                     ILHQKAKRGTPHQAKQAVHCIHAIFTNKEVQLAQIFEPLSRSLNADVPEQLITPLVSL
                     GHISMLAPDQFASPMKSVVANFIVKDLLMNDRSTGEKNGKLWSPDEEVSPEVLAKVQA
                     IKLLVRWLLGMKNNQSKSANSTLRLLSAMLVSEGDLTEQKRISKSDMSRLRLAAGSAI
                     MKLAQEPCYHEIITPEQFQLCALVINDECYQVRQIFAQKLHKALVKLLLPLEYMAIFA
                     LCAKDPVKERRAHARQCLLKNISIRREYIKQNPMATEKLLSLLPEYVVPYMIHLLAHD
                     PDFTRSQDVDQLRDIKECLWFMLEVLMTKNENNSHAFMKKMAENIKLTRDAQSPDESK
                     TNEKLYTVCDVALCVINSKSALCNADSPKDPVLPMKFFTQPEKDFCNDKSYISEETRV
                     LLLTGKPKPAGVLGAVNKPLSATGRKPYVRSTGTETGSNINVNSELNPSTGNRSREQS
                     SEAAETGVSENEENPVRIISVTPVKNIDPVKNKEINSDQATQGNISSDRGKKRTVTAA
                     GAENIQQKTDEKVDESGPPAPSKPRRGRRPKSESQGNATKNDDLNKPINKGRKRAAVG
                     QESPGGLEAGNAKAPKLQDLAKKAAPAERQIDLQR"
BASE COUNT     2156 a   1243 c   1400 g   1945 t
ORIGIN

Alignment Scores:
Pred. No.:              8.71e-243      Length:        6744
Score:                  4760.50        Matches:       935
Percent Similarity:     80.29%         Conservative:  173
Best Local Similarity:  67.75%         Mismatches:    210
Query Match:            66.18%         Indels:        63
DB:                     9              Gaps:          15

US-09-512-581B-2 (1-1391) x AF294791 (1-6744)

Qy          7 ArgThrAsnAspGlyLysIleThrTyrProProGlyValLysGluIleSerAspLysIle 26
              ||||||:::         |||||| ||||||||||||||||||||||||:::|||||||||
Db        160 CGCACACACTCCCGGAAGATCGCTTACCCTCCGGGGGTAAAAGAGATCACCGACAAGATC 219

Qy         27 SerLysGluGluMetValArgArgLeuLysMetValValLysThrPheMetAspMetAsp 46
              :::   :::||||||:::::::||||||||||||||||||||||||||||||||||||||||
Db        220 ACCACGGACGAGATGATCAAACGCCTGAAGATGGTAGTGAAAACCTTTATGGATATGGAT 279

Qy         47 GlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAsp 66
              ||||||||||||:::||||||:::      ||||||  |||||||||||||||||||:::
Db        280 CAGGACTCAGAAGATGAAAAACAGCAGTATCTCCCACTAGCCTTGCATCTTGCATCTGAA 339

Qy         67 PhePheLeuLysHisProGlyLysAspValArgLeuLeuValAlaCysCysLeuAlaAsp 86
              |||||||||:::::|||   ||||||||||||||||||||||||||||||||||||||||
Db        340 TTCTTCCTCAGGAACCCCAATAAAGATGTGCGTCTCCTTGTAGCATGTTGTTTGGCTGAT 399

Qy         87 IlePheArgIleTyrAlaProGluAlaProTyrThrSerProAspLysLeuLysAspIle 106
              ||||||||||||||||||||||||||||||||||||     ||||||||||||||||||
Db        400 ATCTTTCGTATCTATGCCCCAGAAGCTCCATATACTTCCCATGATAAACTTAAGGACATA 459

Qy        107 PheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAsn 126
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 TTTTTGTTTATTACCAGACAATTAAAAGGTTTGGAGGATACAAAGAGTCCACAGTTTAAT 519

Qy        127 ArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsnIleCysPheGlu 146
              |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        520 AGATACTTTTATTTATTAGAGAATTAGCTTGGGTTAAATCATATAACATCTGCTTTGAA 579

Qy        147 LeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsn 166
              |||||||||   |||||||||||     ||||||:::||||||||||||||||||||||||
Db        580 TTGGAAGATTGCAATGAAATTTTTATTCAGCTTTTTAGAACTCTCTTCTCAGTGATCAAC 639

Qy        167 AsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMetSerSerIleIleCys 186
              |||     ||||||:::||||||   |||||||||:::||||||||||||||||||||
Db        640 AATAGCCACAATAAGAAGGTACAAATGCACATGCTAGATTTGATGAGTTCTATCATCATG 699

Qy        187 GluGlyAspThrValSerGlnGluLeuLeuAspThrValLeuValAsnLeuValProAla 206
              |||||||||    |||:::||||||||||||    ::::::|||:::||||||:::||||||
Db        700 GAAGGTGATGGAGTTACTCAAGAATTATTGGGCTCCATTCTTATTAACCTCATTCCTGCA 759

Qy        207 HisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAla 226
              |||||||||||||||||||||:::::::||||||||||||    |||||||||||||||
Db        760 CATAAGAACTTAAATAAACAGTCCTTTGACCTTGCAAAAGTGCTATTGAAAAGAACAGTC 819

Qy        227 GlnAlaIleGluProTyrIleThrThrPhePheAsnGlnValLeuMetLeuGlyLysThr 246
              |||   ||||||     |||        ||||||||||||||||:::||||||::::::
Db        820 CAGACTATTGAGGCATGCATTGCCAATTTTTTCAATCAAGTCCTGGTGCTGGGAAGATCA 879

Qy        247 SerIleSerAspLeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAsp 266
              |||:::||||||||||||||||||||||||||||||||   |||||||:::   ||||||
Db        880 TCAGTAAGTGATTTGTCAGAACATGTATTTGATCTGATTCAGGAACTTTTTGCTATAGAT 939

Qy        267 SerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerAsnAspAsn 286
                 ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db        940 CCTCATTTATTATTATCCGTCATGCCACAGCTTGAATTCAAACTAAAGAGCAATGATGGA 999

Qy        287 GluGluArgLeuGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGlu 306
              ||||||||||||||||||||||:::|||||||||:::|||||||||||||||||||:::
Db       1000 GAAGAGCGATTAGCTGTTGTTCGACTTCTAGCTAAATTGTTTGGCTCCAAAGATTCTGAT 1059

Qy        307 LeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHis 326
              ||||||:::||||||:::|||||||||||||:::||||||||||||||||||||||||||
Db       1060 TTGGCAACACAGAATCGTCCTCTTTGGCAATGTTTTCTTGGACGATTTAATGATATTCAT 1119

Qy        327 ValProIleArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAsp 346
              ||||||:::|||||||||   ||||||||||||||||||||||||||||||||||||||
Db       1120 GTTCCTGTGAGATTAGAAAGTGTGAAATTTGCCAGTCATTGTTTAATGAATCACCCAGAT 1179

Qy        347 LeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIle 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1180 TTAGCGAAGGATCTCACAGAATATTTAAAGGTTAGATCACATGATCCAGAAGAAGCTATT 1239

Qy        367 ArgHisAspValIleValSerIleValThrAlaAlaLysLysAspIleLeuLeuValAsn 386
              |||||||||||||||||||:::|||:::||||||||||||::::::   ||||||||||
Db       1240 CGTCATGATGTCATTGTTACTATAATAACAGCTGCCAAGAGGGACCTGGCCTTAGTAAAT 1299

Qy        387 AspHisLeuLeuAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLys 406
              |||   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db       1300 GATCAGCTGCTTGGCTTTGTAAGGGAAAGAACACTGGATAAACGGTGGCGAGTAAGAAAA 1359

Qy        407 GluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGly 426
              ||||||||||||||||||||||||:::||||||||||||   |||      ||||||
Db       1360 GAAGCTATGATGGGTCTGGCTCAGCTTTATAAGAAATACTGTCTTCATGGTGAAGCAGGA 1419

Qy        427 LysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGln 446
              |||:::||||||::::::::::::||||||||||||||||||||||||||||||||||||
Db       1420 AAGGAAGCTGCAGAGAAAGTCAGCTGGATAAAGGACAAACTTCTGCATATTTATTATCAG 1479

Qy        447 AsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHis 466
              ||||||||||||||||:::||||||||||||:::||||||||||||:::||||||||||||
Db       1480 AACAGCATTGACGACAAACTGTTGGTAGAGAAAATCTTTGCTCAGTATCTTGTCCCCCAC 1539

Qy        467 AsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeu 486
              ||||||||||||   ||||||||||||||||||||||||||||||||:::||||||
Db       1540 AACCTGGAAACAGAAGAGAGAATGAAATGCTTATATTACTTATATGCTAGTTTGGATCCA 1599

Qy        487 AsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnVal 506
              |||||||||||||||||||||||||||||||||||||||||:::||||||      |||
Db       1600 AATGCTGTAAAAGCTCTCAACGAAATGTGGAAGTGTCAGAACATGCTTCGGAGCCATGTA 1659

Qy        507 LysAspLeuLeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePhe 526
              ::::::|||||||||||||   |||||||||   ::::::|||:::       |||:::|||
Db       1660 CGCGAACTATTGGATTTGCACAAGCAGCCTACATCAGAGGCTAACTGTTCTGCCATGTTT 1719

Qy        527 SerLysValMetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMet 546
                 |||:::|||   |||   :::|||||||||||||||||||||||||||||||||:::
```

```
Db   1720 GGAAAACTGATGACCATAGCAAAGAATTTGCCTGACCCCGGGAAAGCACAAGATTTTGTG 1779

Qy    547 LysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeu 566
Db   1780 AAGAAATTTAACCAGGTTCTCGGCGATGATGAGAAACTTCGGTCTCAGTTGGAGTTATTA 1839

Qy    567 ValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLys 586
Db   1840 ATTAGCCCAACCTGTTCTTGCAAACAAGCAGATATTTGTGTGAGAGAAATAGCCCGGAAA 1899

Qy    587 LeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGlu 606
Db   1900 CTTGCAAATCCTAAGCAACCAACAAATCCTTTTCTAGAGATGGTCAAATTTCTGTTGGAA 1959

Qy    607 ArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsn 626
Db   1960 AGAATCGCACCTGTGCACATTGATTCAGAAGCCATAAGTGCACTAGTGAAATTGATGAAT 2019

Qy    627 LysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIle 646
Db   2020 AAGTCAATAGAGGGGACAGCAGATGATGAAGAGGAGGGTGTAAGTCCAGATACAGCTATC 2079

Qy    647 ArgAlaGlyLeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSer 666
Db   2080 CGTTCAGGACTTGAACTTCTTAAGGTTCTGTCTTTTACACATCCTACCTCGTTCCACTCT 2139

Qy    667 AlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGlu 686
Db   2140 GCAGAGACATATGAGTCCTTGTTACAGTGCCTAAGAATGGAGGATGACAAGGTAGCAGAA 2199

Qy    687 AlaAlaLeuGlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHisIle 706
Db   2200 GCTGCTATTCAAATTTTTAGAAATACAGGTCACAAAATAGAAACAGACCTTCCCCAGATA 2259

Qy    707 ArgSerAlaLeuLeuProValLeuHisHisLysSerLysLysGlyProProArgGlnAla 726
Db   2260 CGATCGACCTTAATTCCCATTTTACATCAAAAAGCAAAGAGGGGTACTCCACACCAAGCA 2319

Qy    727 LysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGln 746
Db   2320 AAACAGGCTGTGCACTGTATACACGCCATATTCACAAATAAAGAAGTCCAGCTTGCACAG 2379

Qy    747 IlePheGluProLeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleThrPro 766
Db   2380 ATTTTTGAGCCACTCAGTAGGAGTCTGAATGCTGATGTGCCAGAACAACTTATAACTCCA 2439

Qy    767 LeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLys 786
Db   2440 TTAGTTTCATTGGGCCACATTTCTATGTTAGCACCAGATCAGTTTGCTTCCCCAATGAAA 2499

Qy    787 SerTrpValAlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLys 806
Db   2500 TCTGTAGTAGCAAATTTTATTGTGAAAGATCTGCTAATGAATGACAGGTCAACAGGTGAA 2559

Qy    807 LysThrThrLysLeuTrpValProAspGluGluValSerProGluThrMetValLysIle 826
Db   2560 AAGAATGGAAAACTGTGGTCTCCAGATGAAGAGGTTTCCCCTGAAGTACTAGCAAAGGTA 2619

Qy    827 GlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSer 846
Db   2620 CAGGCAATTAAACTTCTGGTAAGGTGGCTGTTGGGTATGAAAAACAACCAGTCTAAATCT 2679

Qy    847 GlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGlu 866
Db   2680 GCCAATTCAACCCTTCGGTTATTATCAGCGATGTTGGTTAGTGAGGGTGACCTGACAGAG 2739

Qy    867 GlnGlyLysIleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIle 886
Db   2740 CAAAAGAGGATCAGTAAATCTGATATGTCTCGCTTGCGATTAGCTGCTGGTAGTGCCATA 2799

Qy    887 ValLysLeuAlaGlnGluProCysTyrHisGluIleIleThrLeuGluGlnTyrGlnLeu 906
Db   2800 ATGAAGCTTGCTCAGGAACCTTGTTACCATGAAATTATTACCCCAGAACAGTTTCAGCTC 2859

Qy    907 CysAlaLeuAlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeu 926
Db   2860 TGTGCACTTGTTATTAATGATGAGTGTTACCAAGTAAGGCAGATATTTGCTCAGAAGCTG 2919

Qy    927 HisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCys 946
Db   2920 CATAAGGCACTTGTGAAGTTACTGCTCCCATTGGAGTATATGGCGATCTTTGCCTTGTGT 2979

Qy    947 AlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIle 966
Db   2980 GCCAAAGATCCTGTGAAGGAGAGAAGAGCACACGCACGACAATGTTTACTGAAAAATATC 3039

Qy    967 AsnValArgArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeu 986
Db   3040 AGTATACGCAGGGAATACATTAAGCAGAATCCTATGGCTACTGAGAAATTATTATCACTG 3099

Qy    987 LeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrVal 1006
Db   3100 TTGCCTGAATATGTAGTTCCATACATGATTCACCTGCTAGCCCATGATCCAGATTTTACA 3159

Qy   1007 LysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValLeuGlu 1026
Db   3160 AGATCACAAGATGTTGATCAGCTTCGTGATATCAAAGAGTGCCTATGGTTCATGCTTGAA 3219

Qy   1027 IleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsn 1046
Db   3220 GTTTTAATGACAAAGAATGAAAACAATAGCCATGCCTTTATGAAGAAGATGGCAGAGAAC 3279

Qy   1047 IleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyr 1066
Db   3280 ATCAAGTTAACCAGAGATGCCCAGTCTCCAGATGAATCCAAGACAAATGAAAAACTGTAT 3339

Qy   1067 ThrValCysAspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSerLeuGlu 1086
Db   3340 ACAGTATGTGATGTGGCTCTCTGTGTTATAAATAGTAAAAGTGCTTTGTGCAATGCAGAT 3399

Qy   1087 SerProLysAspProValLeuProAlaArgPhePheThrGlnProAspLysAsnPheSer 1106
Db   3400 TCACCAAAGGACCCAGTCCTCCCAATGAAATTTTTACACAACCTGAAAAGGACTTCTGT 3459

Qy   1107 AsnThrLysAsnTyrLeuProProGluMetLysSerPhePheThrProGlyLysProLys 1126
Db   3460 AACGATAAGAGTTATATTTCAGAAGAGACAAGAGTACTTCTGTTAACAGGAAAGCCAAAG 3519

Qy   1127 ThrThrAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGln 1146
Db   3520 CCTGCTGGAGTACTAGGTGCAGTAAATAAGCCTTTATCAGCAACGGGAAGGAAACCCTAT 3579

Qy   1147 ThrLysSerSerArgMetGluThrValSer------AsnAlaSerSerSerSerAsnPro 1164
Db   3580 GTTAGAAGCACTGGCACTGAGACTGGAAGCAATATTAATGTAAATTCAGAGCTGAACCCT 3639

Qy   1165 SerSerProGlyArgIleLysGlyArg---LeuAspSerSerGluMetAspHisSerGlu 1183
Db   3640 TCAACCGGAAATCGATCAAGGGAACAGAGTTCAGAGGCAGCAGAAACTGGAGTTAGTGAA 3699

Qy   1184 AsnGluAsp------TyrThrMetSerSerProLeuProGlyLysLysSerAspLysArg 1201
Db   3700 AATGAAGAGAACCCTGTGAGGATTATTTCAGTCACACCTGTAAAGAATATTGAC------ 3753

Qy   1202 AspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyArgLysLysThrPro 1221
Db   3754 ---------------------------------------------------------CCA 3756

Qy   1222 ValThrGluGlnGluGluLysLeuGlyMetAspAspLeuThrLysLeuValGlnGluGln 1241
Db   3757 GTAAAGAATAAGGAA------ATTAATTCTGATCAGGCTACC---------------CAG 3795

Qy   1242 LysProLysGlySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGlu 1261
Db   3796 GGCAACATCAGCAGTGACCGAGGAAAGAAAAGAACAGTAACAGCAGCTGGTGCAGAGAAT 3855
```

```
Qy       1262 ---GlnGlnTrpProGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGlu 1280
                 ||||||      :::||||||                             ||||||
Db       3856 ATCCAACAAAAAACAGATGAGAAA-------------------------GTAGATGAA 3888

Qy       1281 GlnAsnSerPro---ProLysLysGlyLysArgGlyArgProProLysProLeuGly--- 1298
                       |||   |||   |||   :::|||||||||   ||||||
Db       3889 TCGGGACCTCCCGCCCCTTCCAAACCCAGGAGAGGACGTCGACCCAAGTCTGAATCTCAG 3948

Qy       1299 GlyGlyThrProLysGluGluProThrMetLysThrSerLysLysGlySerLysLysLys 1318
              |||      |||   :::         |||         ||||||   |||:::
Db       3949 GGCAATGCTACCAAAAATGATGATCTAAATAAACCTATTAACAAGGGAAGGAAGAGAGCT 4008

Qy       1319 Ser---GlyProProAlaProGluGluGluGlu------------------------- 1328
              :::   |||      :::|||         |||
Db       4009 GCAGTGGGTCAGGAGAGCCCTGGGGGTTTGGAAGCAGGTAATGCCAAAGCACCCAAACTG 4068

Qy       1329 GluGluGluArgGlnSerGlyAsnThrGluGlnLys------------SerLysSerLys 1344
              ::::::      |||   |||:::|||   :::|||            :::|||   |||
Db       4069 CAAGATTT-AGCCAAAAAGGCAGCACCAGCAGAAAGACAAATTGACTTACAAAGGTAAAA 4127

Qy       1345 GlnHis------ArgValSerArgArgAlaGlnGlnArgAlaGluSerProGluSerSer 1362
                 |||      |||      :::|||:::|||:::   ::::::|||   |||:::
Db       4128 ATGCATTTGCAAAGGGAGAAAATGAAGGCCAAACAGAAG---CAGGCTCCAGCTTCTGCA 4184

RESULT 7
HSU50533
LOCUS       HSU50533              1852 bp    mRNA    linear   PRI 27-NOV-1996
DEFINITION  Human BRCA2 region, mRNA sequence CG008.
ACCESSION   U50533
VERSION     U50533.1  GI:1531605
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1852)
  AUTHORS   Couch,F.J., Rommens,J.M., Neuhausen,S.L., Belanger,C., Dumont,M.,
            Kenneth,A., Bell,R., Berry,S., Bogden,R., Cannon-Albright,L.,
            Farid,L., Frye,C., Hattier,T., Janecki,T., Jiang,P., Kehrer,R.,
            Leblanc,J.-F., McArthur-Morrison,J., McSweeney,D., Miki,Y.,
            Peng,Y., Samson,C., Schroeder,M., Snyder,S.C., Stringfellow,M.,
            Stroup,C., Swedlund,B., Swensen,J., Teng,D., Thakur,S., Tran,T.,
            Tranchant,M., Welver-Feldhaus,J., Wong,A.K.C., Shizuya,H.,
            Labrie,F., Skolnick,M.H., Goldgar,D.E., Kamb,A., Weber,B.L.,
            Tavtigian,S.V. and Simard,J.
  TITLE     Generation of an integrated transcription map of the BRCA2 region
            on chromosome 13q12-q13
  JOURNAL   Genomics 36 (1), 86-99 (1996)
  MEDLINE   96411650
   PUBMED   8812419
REFERENCE   2  (bases 1 to 1852)
  AUTHORS   Simard,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-1996) Jacques Simard, Laboratory of Molecular
            Endocrinology, CHUL Research Center, 2705, Boulevard Laurier,
            Quebec City, Quebec G1V 4G2, Canada
FEATURES             Location/Qualifiers
     source          1..1852
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /map="13q12-q13"
                     /note="CG008; DSEG numbers: D13S171 and D13S310"
BASE COUNT      609 a    348 c    379 g    505 t     11 others
ORIGIN

Alignment Scores:
Pred. No.:              8.93e-148      Length:        1852
Score:                  2953.50        Matches:       589
Percent Similarity:     95.96%         Conservative:  5
Best Local Similarity:  95.15%         Mismatches:    21
Query Match:            41.06%         Indels:        4
DB:                     9              Gaps:          1

US-09-512-581B-2 (1-1391) x HSU50533 (1-1852)

Qy        356 LysValArgSerHisAspProGluGluAlaIleArgHisAspValIleValSerIleVal 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 AAAGTGAGGTCACATGACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTT 62

Qy        376 ThrAlaAlaLysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGlu 395
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 ACAGCTGCTAAAAAGGATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAG 122

Qy        396 ArgThrLeuAspLysArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIle 415
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 AGAACATTAGACAAACGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATT 182

Qy        416 TyrLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrp 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 TATAAGAAATATGCTTTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGG 242

Qy        436 IleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuVal 455
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 ATCAAAGACAAATTGCTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTT 302

Qy        456 GluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLys 475
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 GAACGGATCTTTGCTCAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAA 362

Qy        476 CysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMet 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 TGCTTATATTACTTGTATGCCACACTGGATTTAAATGCTGTGAAAGCATTGAATGAAATG 422

Qy        496 TrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGln 515
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 TGGAAATGTCAAAATCTGCTCCGACATCAAGTAAAGGATTTGCTTGACTTGATTAAGCAA 482

Qy        516 ProLysThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsn 535
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 CCCAAAACAGATGCCAGTGTCAAGGCCATATTTTCAAAAGTGATGGTTATTACAAGAAAT 542

Qy        536 LeuProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAsp 555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 TTACCTGATCCTGGTAAGGCTCAGGATTTCATGAAGAAATTCACACAGGTGTTAGAAGAT 602

Qy        556 AspGluLysIleArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGln 575
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 GATGAGAAAATAAGAAAGCAGTTAGAAGTACTTGTTAGTCCAACATGCTCCTGCAAGCAG 662

Qy        576 AlaGluGlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsn 595
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 GCTGAAGGTTGTGTGCGTGAAATAACTAAGAAGTTGGGCAACCCCAAACAGCCTACAAAT 722

Qy        596 ProPheLeuGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAspThr 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        723 CCTTTCCTGGAAATGATCAAGTTTCTCTTGGAGAGGATAGCACCTGTGCACATAGATACC 782

Qy        616 GluSerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAsp 635
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        783 GAATCTATCAGTGCTCTTATTAAACAAGTGAACAAATCAATAGATGGAACAGCAGATGAT 842

Qy        636 GluAspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysVal 655
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        843 GAAGATGAGGGTGTTCCAACTGATCAAGCCATCAGAGCAGGTCTTGAACTGCTTAAGGTA 902

Qy        656 LeuSerPheThrHisProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAla 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        903 CTCTCATTTACACATCCCATCTCATTTCATTCTGCTGAAACATTTGAATCATTACTGGCT 962

Qy        676 CysLeuLysMetAspAspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThr 695
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        963 TGTCTGAAAATGGATGATGAAAAAGTAGCAGAAGCTGCACTACAAATTTTCAAAAACACA 1022

Qy        696 GlySerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHis 715
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1023 GGAAGCAAAATTGAAGAGGATTTTCCACACATCAGATCAGCCTTGCTTCCTGTTTTACAT 1082

Qy        716 HisLysSerLysLysGlyProProArg-GlnAlaLysTyrAlaIleHisCysIleHisAl 735
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db       1083 CACAAATCTAAAAAAGGACCCCCCCGCTCAAGCCAAATATGCCATTCATTGTATCCATGC 1142

Qy        735 aIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLe 755
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1143 GATATTTTCTAGTAAAGAGACYCAGTTTGCACAGATATTTGAGCCTCTGCATAAGAGCCT 1202

Qy        755 uAspProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLe 775
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1203 AGATCCAAGCAACCTGGAACATCTCATAACACCATTGGTTACTATTGGTCATATTGCTCT 1262

Qy        775 uLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLy 795
              ||||||||||||||||||||||||||  ||||||  ||||||||||||||    ||
Db       1263 CCTTGCACCTGATCAATTTGCTGCTCCTTTGAAATCTTTGGTAGCTACTTTCATTGYGAA 1322

Qy        795 sAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAs 815
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1323 AGATCTTCTCATGAATGATCGGCTTCCAGGGAAAAAGACAACTAAACTTTGGGTTCCAGA 1382

Qy        815 pGluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTr 835
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1383 TGAAGAAGTATCTCCTGAGACAATGGTCAAAATTCAGGCTATTAAAATGATGGTTCGATG 1442

Qy        835 pLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeu-ArgLeuLeuT 855
              |||||||||||||            ||||||  ||||||:::|||:::|||   ||||||
Db       1443 GCTACTTGGCATGTAAMTTAMTCACAGTATATCAGGATCTTCTTCCTTAAGSATTGCTAA 1502

Qy        855 hrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspM 875
              |||||||    |||   |||||||||            :::|||          ||||||
Db       1503 CAACAATATCGCATCGTGATGGGGATTGGMAGAMCAGGGGAAAT---TAGTAACCAGATA 1559

Qy        875 etSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysT 895
              |||||||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db       1560 TGTCACGTCTGAGACTTGCTGCTGGGAGTGCTATTRTGAAGCTGGCACAAGAACCCTGTT 1619

Qy        895 yrHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluC 915
              |||||||||||    |||   |||||||||||||||||||||||||||||||||||||||
Db       1620 ACCATGAAATCRTCACAWTAGAACAATATCAGCTATGTGCRTTAGCTATCAACGATGAAT 1679

Qy        915 ysTyrGlnValArgGlnValPhe-AlaGlnLysLeuHisLysGlyLeuSerArgLeuArg 934
              |||||||||||||||||||||||| ||||||||||||||||  ||||||||||||||||||
Db       1680 GCTATCAAGTAAGACAAGTGTTTGGCCCAGAAACTTCGCAAAGGCCTTTCCCGTTTGCGG 1739

Qy        935 LeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArg 954
              |||||||||||  |||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 CTTCCACTTGAGAATATGGCAATCTGTGCCCTTTGTGCAAAAGATCCTGTAAAGGAGAGA 1799

Qy        955 ArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGlu 971
              |||||||||||||||||||||||||||||||||||||||||||:::|||:::
Db       1800 AGAGCTCATGCTAGGCAATGTTTGGTGAAAAATATAAATGTAAAGCGGGAC 1850
```

```
RESULT 8
AB014548
LOCUS       AB014548                5177 bp    mRNA    linear   PRI 06-FEB-1999
DEFINITION  Homo sapiens mRNA for KIAA0648 protein, partial cds.
ACCESSION   AB014548
VERSION     AB014548.1  GI:3327109
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ishikawa,K., Nagase,T., Suyama,M., Miyajima,N., Tanaka,A.,
            Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes. X.
            The complete sequences of 100 new cDNA clones from brain which can
            code for large proteins in vitro
  JOURNAL   DNA Res. 5 (3), 169-176 (1998)
   MEDLINE  98403880
    PUBMED  9734811
REFERENCE   2  (bases 1 to 5177)
  AUTHORS   Ohara,O., Suyama,M., Nagase,T. and Ishikawa,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-1998) Osamu Ohara, Kazusa DNA Research Institute,
            Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
            292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913,
            Fax:+81-438-52-3914)
FEATURES             Location/Qualifiers
     source          1..5177
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HJ03994"
                     /sex="male"
                     /tissue_type="brain"
                     /clone_lib="pBluescriptII SK plus"
                     /dev_stage="adult"
     gene            1..5177
                     /gene="KIAA0648"
     CDS             <1..2558
                     /gene="KIAA0648"
                     /codon_start=3
                     /product="KIAA0648 protein"
                     /protein_id="BAA31623.1"
                     /db_xref="GI:3327110"
                     /translation="LYYLYASLDPNAVKALNEMWKCQNMLRSHVRELLDLHKQPTSEA
                     NCSAMFGKLMTIAKNLPDPGKAQDFVKKFNQVLGDDEKLRSQLELLISPTCSCKQADI
                     CVREIARKLANPKQPTNPFLEMVKFLLERIAPVHIDSEAISALVKLMNKSIEGTADDE
                     EEGVSPDTAIRSGLELLKVLSFTHPTSFHSAETYESLLQCLRMEDDKVAEAAIQIFRN
                     TGHKIETDLPQIRSTLIPILHQKAKRGTPHQAKQAVHCIHAIFTNKEVQLAQIFEPLS
                     RSLNADVPEQLITPLVSLGHISMLAPDQFASPMKSVVANFIVKDLLMNDRSTGEKNGK
                     LWSPDEEVSPEVLAKVQAIKLLVRWLLGMKNNQSKSANSTLRLLSAMLVSEGDLTEQK
                     RISKSDMSRLRLAAGSAIMKLAQEPCYHEIITPEQFQLCALVINDECYQVRQIFAQKL
                     HKALVKLLLPLEYMAIFALCAKDPVKERRAHARQCLLKNISIRREYIKQNPMATEKLL
                     SLLPEYVVPYMIHLLAHDPDFTRSQDVDQLRDIKECLWFMLEVLMTKNENNSHAFMKK
                     MAENIKLTRDAQSPDESKTNEKLYTVCDVALCVINSKSALCNADSPKDPVLPMKFFTQ
                     PEKDFCNDKSYISEETRVLLLTGKPKPAGVLGAVNKPLSATGRKPYVRSTGTETGSNI
                     NVNSELNPSTGNRSREQSSEAAETGVSENEENPVRIISVTPVKNIDPVKNKEINSDQA
                     TQGNISSDRGKKRTVTAAGAENIQQKTDEKVDESGPPAPSKPRRGRRPKSESQGNATK
                     NDDLNKPINKGRKRAAVGQESPGGLEAGNAKAPKLQDLAKKAAPAERQIDLQR"
BASE COUNT     1690 a    931 c   1037 g   1519 t
ORIGIN

Alignment Scores:
Pred. No.:               5.26e-134      Length:         5177
Score:                   2704.50        Matches:        551
Percent Similarity:      74.07%         Conservative:   123
Best Local Similarity:   60.55%         Mismatches:     174
Query Match:             37.60%         Indels:         63
DB:                      9              Gaps:           15

US-09-512-581B-2 (1-1391) x AB014548 (1-5177)

Qy        477 LeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrp 496
              ||||||||||||||||||:::||||||   ||||||||||||||||||||||||||||||
Db          3 TTATATTACTTATATGCTAGTTTGGATCCAAATGCTGTAAAAGCTCTCAACGAAATGTGG 62

Qy        497 LysCysGlnAsnLeuLeuArgHisGlnValLysAspLeuLeuAspLeuIleLysGlnPro 516
              ||||||||||||:::||||||   |||:::::||||||||||||   ||||||||
Db         63 AAGTGTCAGAACATGCTTCGGAGCCATGTACGCGAACTATTGGATTTGCACAAGCAGCCT 122

Qy        517 LysThrAspAlaSerValLysAlaIlePheSerLysValMetValIleThrArgAsnLeu 536
                 :::::::|||:::     |||:::|||   |||:::|||   |||   :::||||||
Db        123 ACATCAGAGGCTAACTGTTCTGCCATGTTTGGAAAACTGATGACCATAGCAAAGAATTTG 182

Qy        537 ProAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAsp 556
              ||||||||||||||||||||||||||:::|||||||||   |||||||||   ||||||
Db        183 CCTGACCCCGGGAAAGCACAAGATTTTGTGAAGAAATTTAACCAGGTTCTCGGCGATGAT 242

Qy        557 GluLysIleArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAla 576
              ||||||:::|||   |||||||||:::|||:::|||||||||||||||||||||||||||
```

```
Db     243 GAGAAACTTCGGTCTCAGTTGGAGTTATTAATTAGCCCAACCTGTTCTTGCAAACAAGCA 302

Qy     577 GluGlyCysValArgGluIleThrLysLysLeuGlyAsnProLysGlnProThrAsnPro 596
           :::   ||||||||||||||    :::||||||  |||||||||||||||||||||||
Db     303 GATATTTGTGTGAGAGAAATAGCCCGGAAACTTGCAAATCCTAAGCAACCAACAAATCCT 362

Qy     597 PheLeuGluMetIleLysPheLeuLeuGluArgIleAlaProValHisIleAspThrGlu 616
           ||||||||||||:::||||||||||||||||||||||||||||||||||||||:::|||
Db     363 TTTCTAGAGATGGTCAAATTTCTGTTGGAAAGAATCGCACCTGTGCACATTGATTCAGAA 422

Qy     617 SerIleSerAlaLeuIleLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGlu 636
           :::||||||||||||:::|||    :::||||||||||||:::||||||||||||||||||
Db     423 GCCATAAGTGCACTAGTGAAATTGATGAATAAGTCAATAGAGGGGACAGCAGATGATGAA 482

Qy     637 AspGluGlyValProThrAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLysValLeu 656
           :::|||||||||       |||  |||||||||:::||||||||||||||||||||||||
Db     483 GAGGAGGGTGTAAGTCCAGATACAGCTATCCGTTCAGGACTTGAACTTCTTAAGGTTCTG 542

Qy     657 SerPheThrHisProIleSerPheHisSerAlaGluThrPheGluSerLeuLeuAlaCys 676
           |||||||||||||||    ||||||||||||||||||||||:::||||||||||||   |||
Db     543 TCTTTTACACATCCTACCTCGTTCCACTCTGCAGAGACATATGAGTCCTTGTTACAGTGC 602

Qy     677 LeuLysMetAspAspGluLysValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGly 696
           |||:::|||:::|||:::|||||||||||||||||||:::|||||||||:::||||||||
Db     603 CTAAGAATGGAGGATGACAAGGTAGCAGAAGCTGCTATTCAAATTTTAGAAATACAGGT 662

Qy     697 SerLysIleGluGluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHis 716
              |||||||||   |||   |||   |||||||||   |||:::|||:::||||
Db     663 CACAAAATAGAAACAGACCTTCCCCAGATACGATCGACCTTAATTCCCATTTTACATCAA 722

Qy     717 LysSerLysLysGlyProProArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIle 736
           |||:::|||:::|||   |||   |||||||||   |||:::||||||||||||||||||
Db     723 AAAGCAAAGAGGGGTACTCCACACCAAGCAAAACAGGCTGTGCACTGTATACACGCCATA 782

Qy     737 PheSerSerLysGluThrGlnPheAlaGlnIlePheGluProLeuHisLysSerLeuAsp 756
           |||:::::::||||||   |||   ||||||||||||||||||||||   :::|||||||:::
Db     783 TTCACAAATAAAGAAGTCCAGCTTGCACAGATTTTTGAGCCACTCAGTAGGAGTCTGAAT 842

Qy     757 ProSerAsnLeuGluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeu 776
                       |||   ||||||||||||||||||:::::::|||||||||:::::::|||
Db     843 GCTGATGTGCCAGAACAACTTATAACTCCATTAGTTTCATTGGGCCACATTTCTATGTTA 902

Qy     777 AlaProAspGlnPheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAsp 796
           ||||||||||||||||||:::|||   ||||||   |||||   ||||||||||||||||
Db     903 GCACCAGATCAGTTTGCTTCCCCAATGAAATCTGTAGTAGCAAATTTTATTGTGAAAGAT 962

Qy     797 LeuLeuMetAsnAspArgLeuProGlyLysLysThrThrLysLeuTrpValProAspGlu 816
           ||||||||||||||||||      |||:::|||      |||||||||   ||||||||||
Db     963 CTGCTAATGAATGACAGGTCAACAGGTGAAAAGAATGGAAAACTGTGGTCTCCAGATGAA 1022

Qy     817 GluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrpLeu 836
           ||||||||||||||||   :::   |||:::|||||||||||||||:::::::||||||||||||
Db    1023 GAGGTTTCCCCTGAAGTACTAGCAAAGGTACAGGCAATTAAACTTCTGGTAAGGTGGCTG 1082

Qy     837 LeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThrThr 856
           ||||||||||||||||||   |||||||||      ||||||||||||||||||:::
Db    1083 TTGGGTATGAAAAACAACCAGTCTAAATCTGCCAATTCAACCCTTCGGTTATTATCAGCG 1142

Qy     857 IleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMetSer 876
           :::|||   |||:::||||||||||||||||:::|||||||||      |||||||||
Db    1143 ATGTTGGTTAGTGAGGGTGACCTGACAGAGCAAAAGAGGATCAGTAAATCTGATATGTCT 1202

Qy     877 ArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyrHis 896
           ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db    1203 CGCTTGCGATTAGCTGCTGGTAGTGCCATAATGAAGCTTGCTCAGGAACCTTGTTACCAT 1262

Qy     897 GluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsnAspGluCysTyr 916
           ||||||||||||   ||||||:::||||||||||||||||   ||||||||||||||||||
Db    1263 GAAATTATTACCCCAGAACAGTTTCAGCTCTGTGCACTTGTTATTAATGATGAGTGTTAC 1322

Qy     917 GlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeuArgLeuPro 936
           ||||||||||||:::||||||||||||||||||||||   |||   :::|||   ||||||
Db    1323 CAAGTAAGGCAGATATTTGCTCAGAAGCTGCATAAGGCACTTGTGAAGTTACTGCTCCCA 1382

Qy     937 LeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAspProValLysGluArgArgAla 956
           ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db    1383 TTGGAGTATATGGCGATCTTTGCCTTGTGTGCCAAAGATCCTGTGAAGGAGAGAAGAGCA 1442

Qy     957 HisAlaArgGlnCysLeuValLysAsnIleAsnValArgArgGluTyrLeuLysGlnHis 976
           |||||||||||||||||||||||:::|||||||||:::::::||||||||||||:::||||||:::
Db    1443 CACGCACGACAATGTTTACTGAAAAATATCAGTATACGCAGGGAATACATTAAGCAGAAT 1502

Qy     977 AlaAlaValSerGluLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIle 996
                 :::||||||||||||||||||||||||||||||||||||||||||   |||
Db    1503 CCTATGGCTACTGAGAAATTATTATCACTGTTGCCTGAATATGTAGTTCCATACATGATT 1562

Qy     997 HisLeuLeuAlaHisAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAsp 1016
           |||||||||||||||||||||||||||:::   :::   |||||:::::::||||||:::|||
Db    1563 CACCTGCTAGCCCATGATCCAGATTTTACAAGATCACAAGATGTTGATCAGCTTCGTGAT 1622

Qy    1017 ValLysGluCysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSer 1036
           :::||||||||||||||||||:::||||||:::||||||   ||||||||||||||||||
Db    1623 ATCAAAGAGTGCCTATGGTTCATGCTTGAAGTTTAATGACAAAGAATGAAAACAATAGC 1682

Qy    1037 HisAlaPheIleArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyPro 1056
           ||||||||:::::::||||||   ||||||||||||   |||:::||||||||   |||
Db    1683 CATGCCTTTATGAAGAAGATGGCAGAGAACATCAAGTTAACCAGAGATGCCCAGTCTCCA 1742

Qy    1057 AspAspAlaLysMetAsnGluLysLeuTyrThrValCysAspValAlaMetAsnIleIle 1076
           |||:::::::|||   ||||||||||||||||||||||||||||||:::   :::|||
Db    1743 GATGAATCCAAGACAAATGAAAAACTGTATACAGTATGTGATGTGGCTCTCTGTGTTATA 1802

Qy    1077 MetSerLysSerThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArg 1096
              |||||||||         :::   :::||||||||||||||||||||||||   :::
Db    1803 AATAGTAAAAGTGCTTTGTGCAATGCAGATTCACCAAAGGACCCAGTCCTCCCAATGAAA 1862

Qy    1097 PhePheThrGlnProAspLysAsnPheSerAsnThrLysAsnTyrLeuProProGluMet 1116
           ||||||||||||||||:::|||:::|||   |||   |||:::|||:::         |||
Db    1863 TTTTTTACACAACCTGAAAAGGACTTCTGTAACGATAAGAGTTATATTTCAGAAGAGACA 1922

Qy    1117 LysSerPhePheThrProGlyLysProLysThrThrAsnValLeuGlyAlaValAsnLys 1136
           :::                  ||||||||||||         ||||||||||||||||||
Db    1923 AGAGTACTTCTGTTAACAGGAAAGCCAAAGCCTGCTGGAGTACTAGGTGCAGTAAATAAG 1982

Qy    1137 ProLeuSerSerAlaGlyLysGlnSerGlnThrLysSerSerArgMetGluThrValSer 1156
           |||||||||:::   |||:::::::            :::|||:::        ||||||   |||
Db    1983 CCTTTATCAGCAACGGGAAGGAAACCCTATGTTAGAAGCACTGGCACTGAGACTGGAAGC 2042

Qy    1157 ------AsnAlaSerSerSerSerAsnProSerSerProGlyArgIleLysGlyArg--- 1173
                 |||   :::|||         |||||||||:::     |||   :::   :::
Db    2043 AATATTAATGTAAATTCAGAGCTGAACCCTTCAACCGGAAATCGATCAAGGGAACAGAGT 2102

Qy    1174 LeuAspSerSerGluMetAspHisSerGluAsnGluAsp------TyrThrMetSerSer 1191
             :::::::|||         |||||||||||||:::              :::   |||
Db    2103 TCAGAGGCAGCAGAAACTGGAGTTAGTGAAAATGAAGAGAACCCTGTGAGGATTATTTCA 2162

Qy    1192 ProLeuProGlyLysLysSerAspLysArgAspAspSerAspLeuValArgSerGluLeu 1211
              |||   |||   |||
Db    2163 GTCACACCTGTAAAGAATATTGAC------------------------------------ 2186

Qy    1212 GluLysProArgGlyArgLysLysThrProValThrGluGlnGluGluLysLeuGlyMet 1231
                                       ||||||        :::|||         :::
Db    2187 --------------------------CCAGTAAAGAATAAGGAA------ATTAATTCT 2213

Qy    1232 AspAspLeuThrLysLeuValGlnGluGlnLysProLysGlySerGlnArgSerArgLys 1251
           |||      |||               |||            |||   |||   :::|||
Db    2214 GATCAGGCTACC---------------CAGGGCAACATCAGCAGTGACCGAGGAAAGAAA 2258

Qy    1252 ArgGlyHisThrAlaSerGluSerAspGlu---GlnGlnTrpProGluGluLysArgLeu 1270
           |||         ||||||:::   ::::::         ||||||      :::||||||
Db    2259 AGAACAGTAACAGCAGCTGGTGCAGAGAATATCCAACAAAAAACAGATGAGAAA------ 2312

Qy    1271 LysGluAspIleLeuGluAsnGluAspGluGlnAsnSerPro---ProLysLysGlyLys 1289
                                      ||||||      |||   |||   |||   :::
Db    2313 --------------------GTAGATGAATCGGGACCTCCCGCCCCTTCCAAACCCAGG 2351
```

```
Qy      1290 ArgGlyArgProProLysProLeuGly---GlyGlyThrProLysGluGluProThrMet 1308
             ||||||||  ||||||              |||       |||   :::
Db      2352 AGAGGACGTCGACCCAAGTCTGAATCTCAGGGCAATGCTACCAAAAATGATGATCTAAAT 2411

Qy      1309 LysThrSerLysLysGlySerLysLysLysSer---GlyProProAlaProGluGluGlu 1327
             |||      ||||||  |||:::  :::  |||       :::|||
Db      2412 AAACCTATTAACAAGGGAAGGAAGAGAGCTGCAGTGGGTCAGGAGAGCCCTGGGGGTTTG 2471

Qy      1328 Glu--------------------------GluGluGluArgGlnSerGlyAsnThrGlu 1338
             |||                          ::::::    |||  |||:::|||
Db      2472 GAAGCAGGTAATGCCAAAGCACCCAAACTGCAAGATTT-AGCCAAAAAGGCAGCACCAGC 2530

Qy      1339 GlnLys-----------SerLysSerLysGlnHis------ArgValSerArgArgAla 1352
             :::|||           :::|||  |||  |||        |||        :::|||
Db      2531 AGAAAGACAAATTGACTTACAAAGGTAAAAATGCATTTGCAAAGGGAGAAAATGAAGGCC 2590

Qy      1353 GlnGlnArgAlaGluSerProGluSerSer 1362
             :::|||:::   ::::::|||  |||:::
Db      2591 AAACAGAAG---CAGGCTCCAGCTTCTGCA 2617
```

```
RESULT 9
BC039256
LOCUS       BC039256                4252 bp    mRNA    linear   PRI 06-NOV-2002
DEFINITION  Homo sapiens, Similar to androgen-induced prostate proliferative
            shutoff associated protein, clone MGC:33757 IMAGE:5295109, mRNA,
            complete cds.
ACCESSION   BC039256
VERSION     BC039256.1  GI:24657778
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4252)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:      http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 48 Row: k Column: 2
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 7705287.
FEATURES             Location/Qualifiers
     source          1..4252
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:33757 IMAGE:5295109"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
     CDS             71..1660
                     /codon_start=1
                     /product="Similar to androgen-induced prostate
                     proliferative shutoff associated protein"
                     /protein_id="AAH39256.1"
                     /db_xref="GI:24657779"
                     /translation="MAHSKTRTNDGKITYPPGVKEISDKISKEEMVRRLKMVVKTFMD
                     MDQDSEEEKELYLNLALHLASDFFLKHPDKDVRLLVACCLADIFRIYAPEAPYTSPDK
                     LKDIFMFITRQLKGLEDTKSPQFNRYFYLLENIAWVKSYNICFELEDSNEIFTQLYRT
                     LFSVINNGHNQKVHMHMVDLMSSIICEGDTVSQELLDTVLVNLVPAHKNLNKQAYDLA
                     KALLKRTAQAIEPYITNFFNQVLMLGKTSISDLSEHVFDLILELYNIDSHLLLSVLPQ
                     LEFKLKSNDNEERLQVVKLLAKMFGAKDSELASQNKPLWQCYLGRFNDINVPIRLECV
                     KFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVSIVTAAKKDILLVNDHLLNF
                     VRERTLDKRWRVRKEAMMGLAQIYKKYALQSAAGKDAAKQIAWIKDKLLHIYYQNSID
                     DRLLVERIFAQYMVPHNLETTERMKCLYYLYATLDLNAVKYVSNIKFCSFHPLQYIGF
                     YGKETTNTCILKCNLCSVNIV"
BASE COUNT     1407 a    664 c    792 g   1389 t
ORIGIN

Alignment Scores:
Pred. No.:              3.5e-124       Length:          4252
Score:                  2517.00        Matches:         488
Percent Similarity:     98.80%         Conservative:    4
Best Local Similarity:  97.99%         Mismatches:      6
Query Match:            34.99%         Indels:          0
DB:                     9              Gaps:            0

US-09-512-581B-2 (1-1391) x BC039256 (1-4252)

Qy         1 MetAlaHisSerLysThrArgThrAsnAspGlyLysIleThrTyrProProGlyValLys 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        71 ATGGCTCATTCAAAGACTAGGACCAATGATGGAAAAATTACATATCCGCCTGGGGTCAAG 130

Qy        21 GluIleSerAspLysIleSerLysGluGluMetValArgArgLeuLysMetValValLys 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       131 GAAATATCAGATAAAATATCTAAAGAGGAGATGGTGAGACGATTAAAGATGGTTGTGAAA 190

Qy        41 ThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyrLeuAsnLeuAla 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       191 ACTTTTATGGATATGGACCAGGACTCTGAAGAAGAAAAGGAGCTTTATTTAAACCTAGCT 250

Qy        61 LeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspValArgLeuLeuVal 80
             ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db       251 TTACATCTTGCTTCAGATTTTTTTCTCAAGCATCCTGATAAAGATGTTCGCTTACTGGTA 310

Qy        81 AlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaProTyrThrSerPro 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       311 GCCTGCTGCCTTGCTGATATTTTCAGGATTTATGCTCCTGAAGCTCCTTACACATCCCCT 370

Qy       101 AspLysLeuLysAspIlePheMetPheIleThrArgGlnLeuLysGlyLeuGluAspThr 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       371 GATAAACTAAAGGATATATTTATGTTTATAACAAGACAGTTGAAGGGGCTAGAGGATACA 430

Qy       121 LysSerProGlnPheAsnArgTyrPheTyrLeuLeuGluAsnIleAlaTrpValLysSer 140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       431 AAGAGCCCACAATTCAATAGGTATTTTTATTTACTTGAGAACATTGCTTGGGTCAAGTCA 490

Qy       141 TyrAsnIleCysPheGluLeuGluAspSerAsnGluIlePheThrGlnLeuTyrArgThr 160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       491 TATAACATATGCTTTGAGTTAGAAGATAGCAATGAAATTTTCACCCAGCTATACAGAACC 550

Qy       161 LeuPheSerValIleAsnAsnGlyHisAsnGlnLysValHisMetHisMetValAspLeu 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       551 TTATTTTCAGTTATAAACAATGGCCACAATCAGAAAGTCCATATGCACATGGTAGACCTT 610

Qy       181 MetSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeu 200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       611 ATGAGCTCTATTATTTGTGAAGGTGATACAGTGTCTCAGGAGCTTTTGGATACGGTTTTA 670

Qy       201 ValAsnLeuValProAlaHisLysAsnLeuAsnLysGlnAlaTyrAspLeuAlaLysAla 220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       671 GTAAATCTGGTACCTGCTCATAAGAATTTAAACAAGCAAGCATATGATTTGGCAAAGGCT 730

Qy       221 LeuLeuLysArgThrAlaGlnAlaIleGluProTyrIleThrThrPhePheAsnGlnVal 240
             |||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db       731 TTACTGAAGAGGACAGCTCAAGCTATTGAGCCATATATTACCAATTTTTTTAATCAGGTT 790
```

```
Qy        241 LeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHisValPheAspLeuIleLeu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        791 CTGATGCTTGGGAAAACATCTATCAGCGATTTGTCAGAGCATGTCTTTGACTTAATTTTG 850

Qy        261 GluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLeuProGlnLeuGluPheLys 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 GAGCTCTACAATATTGATAGTCATTTGCTGCTCTCTGTTTTACCCCAGCTTGAATTTAAA 910

Qy        281 LeuLysSerAsnAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 TTAAAGAGCAATGATAATGAGGAGCGCCTACAAGTTGTTAAACTACTGGCAAAAATGTTT 970

Qy        301 GlyAlaLysAspSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGly 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        971 GGGGCAAAGGATTCAGAATTGGCTTCTCAAAACAAGCCACTTTGGCAGTGCTACTTGGGC 1030

Qy        321 ArgPheAsnAspIleHisValProIleArgLeuGluCysValLysPheAlaSerHisCys 340
              |||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db       1031 AGGTTTAATGATATCAATGTACCAATCCGCCTGGAATGTGTGAAATTTGCTAGCCATTGT 1090

Qy        341 LeuMetAsnHisProAspLeuAlaLysAspLeuThrGluTyrLeuLysValArgSerHis 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1091 CTCATGAACCATCCTGATTTAGCAAAAGACTTAACAGAGTATCTTAAAGTGAGGTCACAT 1150

Qy        361 AspProGluGluAlaIleArgHisAspValIleValSerIleValThrAlaAlaLysLys 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 GACCCTGAGGAAGCTATTAGACATGATGTTATTGTGTCAATAGTTACAGCTGCTAAAAAG 1210

Qy        381 AspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGluArgThrLeuAspLys 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 GATATTCTTCTGGTCAATGATCACTTACTTAATTTTGTGAGAGAGAGAACATTAGACAAA 1270

Qy        401 ArgTrpArgValArgLysGluAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAla 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1271 CGATGGAGAGTACGCAAAGAAGCCATGATGGGACTTGCCCAAATTTATAAGAAATATGCT 1330

Qy        421 LeuGlnSerAlaAlaGlyLysAspAlaAlaLysGlnIleAlaTrpIleLysAspLysLeu 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1331 TTACAGTCAGCAGCTGGAAAAGATGCTGCAAAACAGATAGCATGGATCAAAGACAAATTG 1390

Qy        441 LeuHisIleTyrTyrGlnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAla 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1391 CTACATATATATTATCAAAATAGTATTGATGATCGACTACTTGTTGAACGGATCTTTGCT 1450

Qy        461 GlnTyrMetValProHisAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeu 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1451 CAATACATGGTTCCTCACAATTTAGAAACTACAGAACGGATGAAATGCTTATATTACTTG 1510

Qy        481 TyrAlaThrLeuAspLeuAsnAlaValLysAlaLeuAsnGluMetTrpLysCys 498
              ||||||||||||||||||||||||||||||   ::::::   :::      |||
Db       1511 TATGCCACACTGGATTTAAATGCTGTGAAGTATGTTTCAAATATCAAATTCTGT 1564
```

RESULT 10
AK098331

```
LOCUS       AK098331                2265 bp    mRNA    linear   PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ41012 fis, clone UTERU2018523, highly similar
            to Human androgen-induced prostate proliferative shutoff associated
            protein (AS3) mRNA.
ACCESSION   AK098331
VERSION     AK098331.1  GI:21758323
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kawakami,B., Sugiyama,A., Takemoto,M., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.
            and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2265)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1..2265
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UTERU2018523"
                     /tissue_type="uterus"
                     /clone_lib="UTERU2"
                     /note="cloning vector: pME18SFL3"
     CDS             125..1927
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAC05286.1"
                     /db_xref="GI:21758324"
                     /translation="MDFTAQPKPATALCGVVSADGKIAYPPGVKEITDKITTDEMIKR
                     LKMVVKTFMDMDQDSEDEKQQYLPLALHLASEFFLRNPNKDVRLLVACCLADIFRIYA
                     PEAPYTSHDKLKDIFLFITRQLKGLEDTKSPQFNRYFYLLENLAWVKSYNICFELEDC
                     NEIFIQLFRTLFSVINNSHNKKVQMHMLDLMSSIIMEGDGVTQELLDSILINLIPAHK
                     NLNKQSFDLAKVLLKRTVQTIEACIANFFNQVLVLGRSSVSDLSEHVFDLIQELFAID
                     PHLLLSVMPQLEFKLKSNDGEERLAVVRLLAKLFGSKDSDLATQNRPLWQCFLGRFND
                     IHVPVRLESVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVTIITAAKRDL
                     ALVNDQLLGFVRERTLDKRWRVRKEAMMGLAQLYKKYCLHGEAGKEAAEKVSWIKDKL
                     LHIYYQNSIDDKLLVEKIFAQYLVPHNLETEERMKCLYYLYASLDPNAVKALNEMWKC
                     QNMLRSHVRELLDLHKQPTSEANCSAMFGKLMTIAKNLPDPGKAQDFVKKFNQVLGDD
                     EKLRSQLELLISPTCSCKQADICVVSKSYFTLFL"
BASE COUNT      696 a    427 c    476 g    666 t
ORIGIN

Alignment Scores:
Pred. No.:              2.94e-122      Length:          2265
Score:                  2474.00        Matches:         460
Percent Similarity:     91.77%         Conservative:    64
Best Local Similarity:  80.56%         Mismatches:      47
Query Match:            34.39%         Indels:          0
DB:                     9              Gaps:            0
```

US-09-512-581B-2 (1-1391) x AK098331 (1-2265)

```
Qy         10 AspGlyLysIleThrTyrProProGlyValLysGluIleSerAspLysIleSerLysGlu 29
              ||||||||||||   ||||||||||||||||||||||||:::|||||||||:::   :::
Db        182 GACGGGAAGATCGCTTACCCTCCGGGGGTAAAAGAGATCACCGACAAGATCACCACGGAC 241

Qy         30 GluMetValArgArgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSer 49
              ||||||::::::||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 GAGATGATCAAACGCCTGAAGATGGTAGTGAAAACCTTTATGGATATGGATCAGGACTCA 301

Qy         50 GluGluGluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPhePheLeu 69
              |||:::||||||:::   ||||||   |||||||||||||||||||||:::|||||||||
Db        302 GAAGATGAAAAACAGCAGTATCTCCCACTAGCCTTGCATCTTGCATCTGAATTCTTCCTC 361

Qy         70 LysHisProGlyLysAspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArg 89
              ::::::|||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 AGGAACCCCAATAAAGATGTGCGTCTCCTTGTAGCATGTTGTTTGGCTGATATCTTTCGT 421

Qy         90 IleTyrAlaProGluAlaProTyrThrSerProAspLysLeuLysAspIlePheMetPhe 109
              ||||||||||||||||||||||||||||||   |||||||||||||||||||||:::|||
Db        422 ATCTATGCCCCAGAAGCTCCATATACTTCCCATGATAAACTTAAGGACATATTTTTGTTT 481
```

```
Qy        110 IleThrArgGlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAsnArgTyrPhe 129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 ATTACCAGACAATTAAAAGGTTTGGAGGATACAAAGAGTCCACAGTTTAATAGATACTTT 541

Qy        130 TyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAsp 149
              |||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        542 TATTTATTAGAGAATTTAGCTTGGGTTAAATCATATAACATCTGCTTTGAATTGGAAGAT 601

Qy        150 SerAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAsnGlyHis 169
                 |||||||||||  ||||||:::||||||||||||||||||||||||||    |||
Db        602 TGCAATGAAATTTTTATTCAGCTTTTTAGAACTCTCTTCTCAGTGATCAACAATAGCCAC 661

Qy        170 AsnGlnLysValHisMetHisMetValAspLeuMetSerSerIleIleCysGluGlyAsp 189
              |||:::||||||  |||||||||:::|||||||||||||||||||||    ||||||||
Db        662 AATAAGAAGGTACAAATGCACATGCTAGATTTGATGAGTTCTATCATCATGGAAGGTGAT 721

Qy        190 ThrValSerGlnGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLysAsn 209
                 |||:::||||||||||||||||:::::|||:::||||||:::|||||||||||||||
Db        722 GGAGTTACTCAAGAATTATTGGACTCCATTCTTATTAACCTCATTCCTGCACATAAGAAC 781

Qy        210 LeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIle 229
              ||||||||||::::::|||||||||||   |||||||||||||||    |||    |||
Db        782 TTAAATAAACAGTCCTTTGACCTTGCAAAAGTGCTATTGAAAAGAACAGTCCAGACTATT 841

Qy        230 GluProTyrIleThrThrPhePheAsnGlnValLeuMetLeuGlyLysThrSerIleSer 249
              |||      |||      |||||||||||||||||:::||||||::::::|||:::|||
Db        842 GAGGCATGCATTGCTAATTTTTCAATCAAGTCCTGGTGCTGGGAAGATCATCAGTAAGT 901

Qy        250 AspLeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeu 269
              |||||||||||||||||||||||||||||   ||||||:::   ||||||   ||||||
Db        902 GATTTGTCAGAACATGTATTTGATCTGATTCAGGAACTTTTGCTATAGATCCTCATTTA 961

Qy        270 LeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluGluArg 289
              |||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db        962 TTATTATCCGTCATGCCACAGCTTGAATTCAAACTAAAGAGCAATGATGGAGAAGAGCGA 1021

Qy        290 LeuGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSer 309
              |||   ||||||:::||||||||||||:::||||||:::|||||||||:::||||||:::
Db       1022 TTAGCTGTTGTTCGACTTCTAGCTAAATTGTTTGGCTCCAAAGATTCTGATTTGGCAACA 1081

Qy        310 GlnAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIle 329
              ||||||:::||||||||||||||||:::||||||||||||||||||||||||||||:::
Db       1082 CAGAATCGTCCTCTTTGGCAATGTTTTCTTGGACGATTTAATGATATTCATGTTCCTGTG 1141

Qy        330 ArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLys 349
              |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1142 AGATTAGAAAGTGTGAAATTTGCCAGTCATTGTTTAATGAATCACCCAGATTTAGCGAAG 1201

Qy        350 AspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleArgHisAsp 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1202 GATCTCACAGAATATTTAAAGGTTAGATCACATGATCCAGAAGAAGCTATTCGTCATGAT 1261

Qy        370 ValIleValSerIleValThrAlaAlaLysLysAspIleLeuLeuValAsnAspHisLeu 389
              |||||||||:::|||:::||||||||||||:::|||:::   |||||||||||    |||
Db       1262 GTCATTGTTACTATAATAACAGCTGCCAAGAGGGACCTGGCCTTAGTAAATGATCAGCTG 1321

Qy        390 LeuAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMet 409
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1322 CTTGGCTTTGTAAGGGAAAGAACACTGGATAAACGGTGGCGAGTAAGAAAAGAAGCTATG 1381

Qy        410 MetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAla 429
              ||||||||||||||:::||||||||||||      |||      |||||||||:::|||
Db       1382 ATGGGTCTGGCTCAGCTTTATAAGAAATACTGTCTTCATGGTGAAGCAGGAAAGGAAGCT 1441

Qy        430 AlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIle 449
              |||::::::::::::|||||||||||||||||||||||||||||||||||||||||||||
Db       1442 GCAGAGAAAGTCAGCTGGATAAAGGACAAACTTCTGCATATTTATTATCAGAACAGCATT 1501

Qy        450 AspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAsnLeuGlu 469
              ||||||:::||||||||||||:::||||||||||||||||:::|||||||||||||||||
Db       1502 GACGACAAACTGTTGGTAGAGAAAATCTTTGCTCAGTATCTTGTCCCCCACAACCTGGAA 1561

Qy        470 ThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaVal 489
              |||   |||||||||||||||||||||||||||||||:::||||||    |||||||||
Db       1562 ACAGAAGAGAGAATGAAATGCTTATATTACTTATATGCTAGTTTGGATCCAAATGCTGTA 1621

Qy        490 LysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeu 509
              |||||||||||||||||||||||||||||||||:::||||||    |||::::::|||
Db       1622 AAAGCTCTCAACGAAATGTGGAAGTGTCAGAACATGCTTCGGAGCCATGTACGCGAACTA 1681

Qy        510 LeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysVal 529
              |||||||||   |||||||||   ::::::|||:::      |||:::|||   |||:::
Db       1682 TTGGATTTGCACAAGCAGCCTACATCAGAGGCTAACTGTTCTGCCATGTTTGGAAAACTG 1741

Qy        530 MetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPhe 549
              |||   |||   :::|||||||||||||||||||||||||||||||||:::|||||||||
Db       1742 ATGACCATAGCAAAGAATTTGCCTGACCCCGGGAAAGCACAAGATTTTGTGAAGAAATTT 1801

Qy        550 ThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerPro 569
                 |||||||||   ||||||||||||:::|||   |||||||||:::|||:::||||||
Db       1802 AACCAGGTTCTCGGCGATGATGAGAAACTTCGGTCTCAGTTGGAGTTATTAATTAGCCCA 1861

Qy        570 ThrCysSerCysLysGlnAlaGluGlyCysVal 580
              |||||||||||||||||||||:::   ||||||
Db       1862 ACCTGTTCTTGCAAACAAGCAGATATTTGTGTG 1894
```

```
RESULT 11
BC041361
LOCUS       BC041361                2705 bp    mRNA    linear   PRI 09-JAN-2003
DEFINITION  Homo sapiens, Similar to KIAA0648 protein, clone MGC:43818
            IMAGE:5273075, mRNA, complete cds.
ACCESSION   BC041361
VERSION     BC041361.1  GI:27552793
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2705)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 75 Row: c Column: 19.
FEATURES             Location/Qualifiers
     source          1..2705
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:43818 IMAGE:5273075"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
```

```
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
     CDS             545..2347
                     /codon_start=1
                     /product="Similar to KIAA0648 protein"
                     /protein_id="AAH41361.1"
                     /db_xref="GI:27552794"
                     /translation="MDFTAQPKPATALCGVVSADGKIAYPPGVKEITDKITTDEMIKR
                     LKMVVKTFMDMDQDSEDEKQQYLPLALHLASEFFLRNPNKDVRLLVACCLADIFRIYA
                     PEAPYTSHDKLKDIFLFITRQLKGLEDTKSPQFNRYFYLLENLAWVKSYNICFELEDC
                     NEIFIQLFRTLFSVINNSHNKKVQMHMLDLMSSIIMEGDGVTQELLDSILINLIPAHK
                     NLNKQSFDLAKVLLKRTVQTIEACIANFFNQVLVLGRSSVSDLSEHVFDLIQELFAID
                     PHLLLSVMPQLEFKLKSNDGEERLAVVRLLAKLFGSKDSDLATQNRPLWQCFLGRFND
                     IHVPVRLESVKFASHCLMNHPDLAKDLTEYLKVRSHDPEEAIRHDVIVTIITAAKRDL
                     ALVNDQLLGFVRERTLDKRWRVRKEAMMGLAQLYKKYCLHGEAGKEAAEKVSWIKDKL
                     LHIYYQNSIDDKLLVEKIFAQYLVPHNLETEERMKCLYYLYASLDPNAVKALNEMWKC
                     QNMLRSHVRELLDLHKQPTSEANCSAMFGKLMTIAKNLPDPGKAQDFVKKFNQVLGDD
                     EKLRSQLELLISPTCSCKQADICVVSKSYFTLFL"
BASE COUNT     751 a    589 c    654 g    711 t
ORIGIN

Alignment Scores:
Pred. No.:                3.7e-122      Length:          2705
Score:                    2474.00       Matches:         460
Percent Similarity:       91.77%        Conservative:    64
Best Local Similarity:    80.56%        Mismatches:      47
Query Match:              34.39%        Indels:          0
DB:                       9             Gaps:            0

US-09-512-581B-2 (1-1391) x BC041361 (1-2705)

Qy         10 AspGlyLysIleThrTyrProProGlyValLysGluIleSerAspLysIleSerLysGlu 29
              |||||||||||    |||||||||||||||||||||||:::|||||||||:::   :::
Db        602 GACGGGAAGATCGCTTACCCTCCGGGGGTAAAAGAGATCACCGACAAGATCACCACGGAC 661

Qy         30 GluMetValArgArgLeuLysMetValValLysThrPheMetAspMetAspGlnAspSer 49
              ||||||:::::|||||||||||||||||||||||||||||||||||||||||||||||||
Db        662 GAGATGATCAAACGCCTGAAGATGGTAGTGAAAACCTTTATGGATATGGATCAGGACTCA 721

Qy         50 GluGluGluLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPhePheLeu 69
              |||:::||||||:::     ||||||    ||||||||||||||||||:::|||||||||
Db        722 GAAGATGAAAAACAGCAGTATCTCCCACTAGCCTTGCATCTTGCATCTGAATTCTTCCTC 781

Qy         70 LysHisProGlyLysAspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArg 89
              ::::::|||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 AGGAACCCCAATAAAGATGTGCGTCTCCTTGTAGCATGTTGTTTGGCTGATATCTTTCGT 841

Qy         90 IleTyrAlaProGluAlaProTyrThrSerProAspLysLeuLysAspIlePheMetPhe 109
              ||||||||||||||||||||||||||||   |||||||||||||||||||||:::|||
Db        842 ATCTATGCCCCAGAAGCTCCATATACTTCCCATGATAAACTTAAGGACATATTTTTGTTT 901

Qy        110 IleThrArgGlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAsnArgTyrPhe 129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 ATTACCAGACAATTAAAAGGTTTGGAGGATACAAAGAGTCCACAGTTTAATAGATACTTT 961

Qy        130 TyrLeuLeuGluAsnIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAsp 149
              ||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||||
Db        962 TATTTATTAGAGAATTTAGCTTGGGTTAAATCATATAACATCTGCTTTGAATTGGAAGAT 1021

Qy        150 SerAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsnAsnGlyHis 169
                 |||||||||||    ||||||:::||||||||||||||||||||||||||   |||
Db       1022 TGCAATGAAATTTTTATTCAGCTTTTTAGAACTCTCTTCTCAGTGATCAACAATAGCCAC 1081

Qy        170 AsnGlnLysValHisMetHisMetValAspLeuMetSerSerIleIleCysGluGlyAsp 189
              |||:::||||||   |||||||||:::|||||||||||||||||||   |||||||||
Db       1082 AATAAGAAGGTACAAATGCACATGCTAGATTTGATGAGTTCTATCATCATGGAAGGTGAT 1141

Qy        190 ThrValSerGlnGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLysAsn 209
                 |||:::|||||||||||||||:::::|||:::||||||:::|||||||||||||||
Db       1142 GGAGTTACTCAAGAATTATTGGACTCCATTCTTATTAACCTCATTCCTGCACATAAGAAC 1201

Qy        210 LeuAsnLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIle 229
              ||||||||||||::::::||||||||||||    |||||||||||||||   |||   |||
Db       1202 TTAAATAAACAGTCCTTTGACCTTGCAAAAGTGCTATTGAAAAGAACAGTCCAGACTATT 1261

Qy        230 GluProTyrIleThrThrPhePheAsnGlnValLeuMetLeuGlyLysThrSerIleSer 249
              |||      |||      ||||||||||||||||||:::||||||::::::|||:::|||
Db       1262 GAGGCATGCATTGCTAATTTTTTCAATCAAGTCCTGGTGCTGGGAAGATCATCAGTAAGT 1321

Qy        250 AspLeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeu 269
              |||||||||||||||||||||||||||      ||||||:::   ||||||   ||||||
Db       1322 GATTTGTCAGAACATGTATTTGATCTGATTCAGGAACTTTTTGCTATAGATCCTCATTTA 1381

Qy        270 LeuLeuSerValLeuProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluGluArg 289
              |||||||||:::|||||||||||||||||||||||||||||||||||   |||||||||
Db       1382 TTATTATCCGTCATGCCACAGCTTGAATTCAAACTAAAGAGCAATGATGGAGAAGAGCGA 1441

Qy        290 LeuGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaSer 309
              |||   |||||:::||||||||||||:::|||||||||:::||||||||||||:::|||:::
Db       1442 TTAGCTGTTGTTCGACTTCTAGCTAAATTGTTTGGCTCCAAAGATTCTGATTTGGCAACA 1501

Qy        310 GlnAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProIle 329
              ||||||:::|||||||||||||||:::||||||||||||||||||||||||||||||:::
Db       1502 CAGAATCGTCCTCTTTGGCAATGTTTTCTTGGACGATTTAATGATATTCATGTTCCTGTG 1561

Qy        330 ArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLys 349
              |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1562 AGATTAGAAAGTGTGAAATTTGCCAGTCATTGTTTAATGAATCACCCAGATTTAGCGAAG 1621

Qy        350 AspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleArgHisAsp 369
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1622 GATCTCACAGAATATTTAAAGGTTAGATCACATGATCCAGAAGAAGCTATTCGTCATGAT 1681

Qy        370 ValIleValSerIleValThrAlaAlaLysLysAspIleLeuLeuValAsnAspHisLeu 389
              |||||||||:::|||:::|||||||||||||:::|||:::   |||||||||||   |||
Db       1682 GTCATTGTTACTATAATAACAGCTGCCAAGAGGGACCTGGCCTTAGTAAATGATCAGCTG 1741

Qy        390 LeuAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLysGluAlaMet 409
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1742 CTTGGCTTTGTAAGGGAAAGAACACTGGATAAACGGTGGCGAGTAAGAAAAGAAGCTATG 1801

Qy        410 MetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAla 429
              ||||||||||||||:::|||||||||||   |||      ||||||||||:::|||
Db       1802 ATGGGTCTGGCTCAGCTTTATAAGAAATACTGTCTTCATGGTGAAGCAGGAAAGGAAGCT 1861

Qy        430 AlaLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIle 449
              |||::::::::::|||||||||||||||||||||||||||||||||||||||||||||||
Db       1862 GCAGAGAAAGTCAGCTGGATAAAGGACAAACTTCTGCATATTTATTATCAGAACAGCATT 1921

Qy        450 AspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAsnLeuGlu 469
              ||||||:::||||||||||||:::||||||||||||:::|||||||||||||||||||||
Db       1922 GACGACAAACTGTTGGTAGAGAAAATCTTTGCTCAGTATCTTGTCCCCCACAACCTGGAA 1981

Qy        470 ThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaVal 489
              |||   ||||||||||||||||||||||||||||||:::||||||   |||||||||
Db       1982 ACAGAAGAGAGAATGAAATGCTTATATTACTTATATGCTAGTTTGGATCCAAATGCTGTA 2041

Qy        490 LysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeu 509
              ||||||||||||||||||||||||||||||||||||:::||||||      |||:::::|||
Db       2042 AAAGCTCTCAACGAAATGTGGAAGTGTCAGAACATGCTTCGGAGCCATGTACGCGAACTA 2101

Qy        510 LeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysVal 529
              |||||||||   |||||||||   ::::::|||:::   |||:::|||   |||:::
Db       2102 TTGGATTTGCACAAGCAGCCTACATCAGAGGCTAACTGTTCTGCCATGTTTGGAAAACTG 2161

Qy        530 MetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPhe 549
              |||   |||   :::|||||||||||||||||||||||||||||||||:::::|||||||
Db       2162 ATGACCATAGCAAAGAATTTGCCTGACCCCGGGAAAGCACAAGATTTTGTGAAGAAATTT 2221

Qy        550 ThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerPro 569
                 |||||||||   ||||||||||||:::|||   |||||||||:::|||:::||||||
Db       2222 AACCAGGTTCTCGGCGATGATGAGAAACTTCGGTCTCAGTTGGAGTTATTAATTAGCCCA 2281

Qy        570 ThrCysSerCysLysGlnAlaGluGlyCysVal 580
              ||||||||||||||||||:::   ||||||
Db       2282 ACCTGTTCTTGCAAACAAGCAGATATTTGTGTG 2314
```

```
RESULT 12
BC021408
LOCUS       BC021408              1976 bp    mRNA    linear   ROD 16-APR-2003
DEFINITION  Mus musculus RIKEN cDNA 9030416H16 gene, mRNA (cDNA clone MGC:29439
            IMAGE:3964501), complete cds.
ACCESSION   BC021408
VERSION     BC021408.1  GI:18204099
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1976)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1976)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 39 Row: b Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: GenomeScan gene
            prediction.
FEATURES             Location/Qualifiers
     source          1. .1976
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="MGC:29439 IMAGE:3964501"
                     /tissue_type="Mammary tumor. Metallothionien-TGF alpha
                     model. 10 month old virgin mouse. Taken by biopsy."
                     /clone_lib="NCI_CGAP_Mam1"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1. .1976
                     /gene="9030416H16Rik"
                     /db_xref="LocusID:71521"
                     /db_xref="MGI:1918771"
     CDS             130. .1884
                     /codon_start=1
                     /product="RIKEN cDNA 9030416H16"
                     /protein_id="AAH21408.1"
                     /db_xref="GI:18204100"
                     /db_xref="LocusID:71521"
                     /translation="MFGKLMTIAKNLPDPGKAQDFVKKFNQVLGDDEKLRSQLELLIS
                     PTCSCKQADVCVREIARKLANPKQPTNPFLEMVKFLLERIAPVHIDSEAISALVKLMN
                     KSIEGTADDEEGVSPDSAIRSGLELLKVLSFTHPTSFHSAETYESLLQCLRMEDDKV
                     AEAAIQIFRNTGHKIETDLPQIRSTLIPILHQKAKRGTPHQAKQAVHCIHAIFSNKEV
                     QLAQIFEPLSRSLNADVPEQLITPLVSLGHISMLAPDQFASPMKSVVANFIVKDLLMN
                     DRSTGEKNGKLWSPDEEVSPEVLAKVQAIKLLVRWLLGMKNNQSKSANSTLRLLSAML
                     VSEGDLTEQKRISKSDMSRLRLAAGSAIMKLAQEPCYHEIITPEQFQLCALVINDECY
                     QVRQIFAQKLHKALVKLLLPLEYMAIFALCAKDPVKERRAHARQCLLKNISIRREYIK
                     QNPMATEKLLSLLPEYVVPYMIHLLAHDPDFTRSQDVDQLRDIKECLWFMLEVLMTKN
                     ENNSHAFMKKMAENIKLTRDAQSPDESKTNEKLYTVCDVALCVINSKSALCNADSPKD
                     PVLPMKFFTQPEKVISLT"
BASE COUNT      611 a    429 c    432 g    504 t
ORIGIN

Alignment Scores:
Pred. No.:              1.22e-120      Length:        1976
Score:                  2442.00        Matches:       457
Percent Similarity:     86.75%         Conservative:  80
Best Local Similarity:  73.83%         Mismatches:    82
Query Match:            33.95%         Indels:        0
DB:                     10             Gaps:          0

US-09-512-581B-2 (1-1391) x BC021408 (1-1976)

Qy        490 LysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysAspLeu 509
              ||||||||||||||||||||||||||||||||||||:::||||||      |||::::::|||
Db         25 AAAGCTCTCAATGAAATGTGGAAATGTCAGAACATGCTTCGAAGTCATGTGCGTGAACTA 84

Qy        510 LeuAspLeuIleLysGlnProLysThrAspAlaSerValLysAlaIlePheSerLysVal 529
              |||||||||   |||||||||   ::::::|||:::      |||:::|||   |||:::
Db         85 TTGGACTTACACAAGCAGCCTACATCAGAAGCGAACTGTTCTGCCATGTTTGGGAAACTG 144

Qy        530 MetValIleThrArgAsnLeuProAspProGlyLysAlaGlnAspPheMetLysLysPhe 549
              |||   |||   :::|||||||||||||||||||||||||||||||||:::|||||||||
Db        145 ATGACCATAGCAAAGAATTTGCCTGACCCTGGAAAAGCACAAGATTTTGTAAAGAAATTT 204

Qy        550 ThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeuGluValLeuValSerPro 569
                 |||||||||   ||||||||||||:::|||   |||||||||:::|||:::||||||
Db        205 AACCAGGTTCTTGGTGATGATGAGAAACTGAGGTCTCAGTTGGAATTATTAATCAGCCCA 264

Qy        570 ThrCysSerCysLysGlnAlaGluGlyCysValArgGluIleThrLysLysLeuGlyAsn 589
              |||||||||||||||||||||||:::   |||||||||||||||   :::||||||   |||
Db        265 ACCTGTTCATGCAAGCAGGCTGACGTTTGTGTGAGGGAAATAGCTCGGAAACTTGCAAAT 324

Qy        590 ProLysGlnProThrAsnProPheLeuGluMetIleLysPheLeuLeuGluArgIleAla 609
              |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        325 CCTAAGCAGCCAACCAACCCTTTTCTAGAGATGGTCAAATTTCTTTTGGAAAGAATTGCT 384

Qy        610 ProValHisIleAspThrGluSerIleSerAlaLeuIleLysGlnValAsnLysSerIle 629
              |||||||||||||||:::|||:::||||||||||||:::|||   :::||||||||||||
Db        385 CCTGTGCACATCGATTCAGAAGCCATAAGTGCACTGGTAAAACTGATGAATAAGTCAATC 444

Qy        630 AspGlyThrAlaAspAspGluAspGluGlyValProThrAspGlnAlaIleArgAlaGly 649
              :::||||||||||||||||||:::|||||||||      |||   |||||||||:::|||
```

```
Db        445 GAAGGGACGGCCGATGATGAAGAGGAGGGTGTCAGTCCAGACTCAGCCATCCGCTCAGGA 504

Qy        650 LeuGluLeuLeuLysValLeuSerPheThrHisProIleSerPheHisSerAlaGluThr 669
Db        505 CTTGAGCTTCTTAAGGTTCTGTCTTTCACACATCCTACCTCGTTCCACTCTGCAGAGACA 564

Qy        670 PheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGluAlaAlaLeu 689
Db        565 TATGAGTCCTTGTTACAGTGCCTAAGAATGGAGGATGACAAGGTAGCAGAAGCTGCAATA 624

Qy        690 GlnIlePheLysAsnThrGlySerLysIleGluGluAspPheProHisIleArgSerAla 709
Db        625 CAAATTTTTAGAAACACAGGCCACAAAATAGAAACTGACCTTCCCCAGATACGGTCAACC 684

Qy        710 LeuLeuProValLeuHisHisLysSerLysLysGlyProProArgGlnAlaLysTyrAla 729
Db        685 TTGATCCCCATTTTACATCAGAAAGCCAAGCGGGGCACTCCACACCAAGCAAAGCAGGCT 744

Qy        730 IleHisCysIleHisAlaIlePheSerSerLysGluThrGlnPheAlaGlnIlePheGlu 749
Db        745 GTTCACTGCATCCATGCCATCTTCTCAAACAAGGAGGTCCAGCTGGCACAGATTTTTGAG 804

Qy        750 ProLeuHisLysSerLeuAspProSerAsnLeuGluHisLeuIleThrProLeuValThr 769
Db        805 CCACTCAGTAGGAGTCTGAATGCTGATGTACCAGAGCAACTTATAACTCCATTAGTTTCA 864

Qy        770 IleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAlaProTrpLysSerTrpVal 789
Db        865 CTGGGCCACATTTCCATGTTAGCACCAGATCAGTTTGCCTCCCCGATGAAATCTGTAGTG 924

Qy        790 AlaThrPheIleValLysAspLeuLeuMetAsnAspArgLeuProGlyLysLysThrThr 809
Db        925 GCAAACTTTATTGTTAAAGACCTTCTAATGAACGACAGGTCAACAGGTGAGAAGAATGGA 984

Qy        810 LysLeuTrpValProAspGluGluValSerProGluThrMetValLysIleGlnAlaIle 829
Db        985 AAATTATGGTCTCCAGATGAGGAAGTTTCTCCTGAAGTTCTAGCAAAGGTACAGGCAATT 1044

Qy        830 LysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSer 849
Db       1045 AAACTTCTGGTAAGGTGGCTGTTGGGTATGAAAAACAACCAGTCTAAATCTGCCAACTCA 1104

Qy        850 ThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLys 869
Db       1105 ACTCTTCGGTTATTATCAGCCATGCTGGTTAGTGAGGGCGACCTGACAGAGCAAAAGAGG 1164

Qy        870 IleSerLysProAspMetSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeu 889
Db       1165 ATCAGTAAATCTGATATGTCTCGCTTGCGATTAGCTGCTGGTAGTGCCATAATGAAGCTT 1224

Qy        890 AlaGlnGluProCysTyrHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeu 909
Db       1225 GCTCAGGAACCTTGCTACCATGAAATTATTACCCCAGAACAGTTTCAGCTCTGTGCACTG 1284

Qy        910 AlaIleAsnAspGluCysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGly 929
Db       1285 GTTATTAATGATGAGTGCTACCAAGTAAGGCAGATATTTGCCCAGAAGCTTCATAAGGCA 1344

Qy        930 LeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCysAlaLysAsp 949
Db       1345 CTTGTGAAGTTGCTTCTCCCACTGGAGTATATGGCCATCTTTGCTTTGTGTGCCAAAGAC 1404

Qy        950 ProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIleAsnValArg 969
Db       1405 CCTGTGAAGGAAAGGAGAGCACACGCTCGGCAGTGTTTGTTAAAGAACATCAGCATCCGC 1464

Qy        970 ArgGluTyrLeuLysGlnHisAlaAlaValSerGluLysLeuLeuSerLeuLeuProGlu 989
Db       1465 AGGGAGTACATCAAACAGAACCCCATGGCCACTGAGAAATTATTGTCACTGCTGCCTGAA 1524

Qy        990 TyrValValProTyrThrIleHisLeuLeuAlaHisAspProAspTyrValLysValGln 1009
Db       1525 TATGTGGTTCCATACATGATTCACCTCCTAGCCCATGATCCTGATTTCACACGGTCACAA 1584

Qy       1010 AspIleGluGlnLeuLysAspValLysGluCysLeuTrpPheValLeuGluIleLeuMet 1029
Db       1585 GATGTTGATCAACTTCGTGATATAAAAGAGTGCCTGTGGTTTATGCTTGAAGTCTTAATG 1644

Qy       1030 AlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysMetValGluAsnIleLysGln 1049
Db       1645 ACAAAGAATGAAAACAACAGCCATGCATTCATGAAAAAGATGGCAGAGAATATCAAGCTA 1704

Qy       1050 ThrLysAspAlaGlnGlyProAspAspAlaLysMetAsnGluLysLeuTyrThrValCys 1069
Db       1705 ACCAGAGACGCCCAGTCTCCCGATGAATCCAAGACAAATGAAAAACTTTATACGGTTGT 1764

Qy       1070 AspValAlaMetAsnIleIleMetSerLysSerThrThrTyrSerLeuGluSerProLys 1089
Db       1765 GACGTGGCTCTGTGTGTTATAAATAGTAAAAGTGCTTTGTGCAATGCAGACTCACCAAAG 1824

Qy       1090 AspProValLeuProAlaArgPhePheThrGlnProAspLysAsnPheSerAsnThr 1108
Db       1825 GACCCAGTCCTCCCAATGAAGTTCTTTACACAGCCTGAAAAGGTAATTTCCCTCACA 1881
```

```
RESULT 13
AL844559
LOCUS       AL844559               216608 bp    DNA     linear   VRT 28-JAN-2003
DEFINITION  Zebrafish DNA sequence from clone DKEY-14K1, complete sequence.
ACCESSION   AL844559
VERSION     AL844559.14  GI:28144648
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 216608)
  AUTHORS   Babbage,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 30, 2003 this sequence version replaced gi:27801645.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Repeat names
            beginning 'Dr' were identified by the Recon repeat discovery system
            (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr'
            were identified by Rick Waterman (Stephen Johnson lab, WashU).  For
            further information see http://www/Projects/D_rerio/fishmask.shtml
            DKEY-14K1 is from a Zebrafish BAC library
            VECTOR: pIndigoBAC-5.
FEATURES             Location/Qualifiers
     source          1..216608
```

```
                    /organism="Danio rerio"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:7955"
                    /clone="DKEY-14K1"
                    /clone_lib="DanioKey"
BASE COUNT     68345 a  39077 c  40063 g  69123 t
ORIGIN

Alignment Scores:
Pred. No.:              6.79e-86       Length:         216608
Score:                  1836.00        Matches:        916
Percent Similarity:     18.40%         Conservative:   148
Best Local Similarity:  15.84%         Mismatches:     215
Query Match:            25.52%         Indels:         4511
DB:                     5              Gaps:           39

US-09-512-581B-2 (1-1391) x AL844559 (1-216608)

Qy          37 MetValValLysThrPheMetAspMetAspGlnAspSerGluGluGluLysGluLeuTyr 56
               :::||||||||||:::||||||||||||||||||||||||||||||||:::    |||
Db       95986 TTGGTGGTTAAGACATATATGGACATGGATCAAGACTCAGAAGAGGAAAAGCAACAGTAC 96045

Qy          57 LeuAsnLeuAlaLeuHisLeuAlaSerAspPhePheLeuLysHisProGlyLysAspVal 76
               |||    ||||||||||||||||||||:::|||||||||:::::|||    ||||||||
Db       96046 TTGGCTTTGGCTCTTCACCTGGCCTCTGAGTTCTTTCTGCGAAACCCCAACAAGGATGTA 96105

Qy          77 ArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAlaProGluAlaPro 96
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       96106 CGGTTACTCGTGGCTTGCTGTCTCGCTGACATCTTCAGAATCTATGCTCCAGAGGCACCA 96165

Qy          97 TyrThrSerProAspLysLeuLys------------------------------------ 104
               |||||||||    |||||||||||
Db       96166 TACACCTCACATGATAAACTCAAGGTATAACCAATACTCATCATCTAGCTGTCTTGAGAA 96225

Qy         104 ------------------------------------------------------------ 104

Db       96226 ATTGTTAAATCTTTCAAAAACACTTTTACGTTGTCAAACTTGATTTCTTTCTACCTGCAC 96285

Qy         104 ------------------------------------------------------------ 104

Db       96286 ACAATCACCTCACTCATACGAAACTAAATTGGTATTATTTTCAGAGTTTAAAACGAGTAC 96345

Qy         105 ---------------------------------AspIlePheMetPheIleThrArgGln 113
                                                :::|||||||:::||||||||||||||
Db       96346 AGTGATGATGGTGATTTGTTTTTCCCTATCTAGGAAATTTTTCTTTTCATCACACGGCAG 96405

Qy         114 LeuLysGlyLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLeuGlu 133
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       96406 TTGAAGGGTCTTGAGGACACCAAGAGCCCACAGTTCAACAGATATTTTTACTTGCTAGAG 96465

Qy         133 ------------------------------------------------------------ 133

Db       96466 GTGAGATGGCTCTTTTAATGAGAGTTTATTGATCTTTTGTCCTCTGTTTAGACAAAGAAA 96525

Qy         133 ------------------------------------------------------------ 133

Db       96526 AGGCTTTGAATGCTTGTTTTAAATTTACTCTCAGACTCAAGCCATCTAGCAGCTTTTTTG 96585

Qy         133 ------------------------------------------------------------ 133

Db       96586 ACCAGTTTGCCTTTGGTTTTAGATCGTATTTATGATTGTCATGACAGCTAACTTTTCAAT 96645

Qy         134 -------------AsnIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAs 149
                            |||:::|||||||||||||||||||||||||||||||||||||||||
Db       96646 TTGACTTGCACAGAACTTGGCATGGGTGAAGTCCTACAACATCTGTTTTGAATTGGAGGA 96705

Qy         149 pSerAsnGluIlePheThrGlnLeuTyrArgThrLeuPheSerValIleAsn-------- 166
               |    |||||||||||    ||||||:::::::||||||||||||||||
Db       96706 CTGTAATGAAATCTTTATCCAGCTCTTCAAAACCCTCTTTTCTGTCATCAA-GTGAGTTA 96764

Qy         166 ------------------------------------------------------------ 166

Db       96765 ATTTGAAGTCAGTGTAATTATCCCATTTGTTCCTCAACTGTTCTCTTTCTGATCAGCAC 96824

Qy         167 ----------------AsnGlyHisAsnGlnLysValHisMetHisMetValAspLeuMe 181
                               |||    ||||||||||||||    |||||||||:::|||||||
Db       96825 AATGCACTCTTTCAGTAACAGCCATAATCAGAAGGTGCAAATGCACATGTTGGACCTGAT 96884

Qy         181 tSerSerIleIleCysGluGlyAspThrValSerGlnGluLeuLeuAspThrValLeuVa 201
               ||||||||||||    ||||||||    |||:::||||||||||||||||||:::|||::
Db       96885 GAGTTCCATCATCATGGAGGGTGATGGAGTGACGCAAGAGCTGCTGGACACAATCCTCAT 96944

Qy         201 lAsnLeuValProAlaHisLys-------------------------------------- 208
               :||||||:::|||||||||||
Db       96945 CAACCTCATCCCTGCTCACAAAGTAGGCAAATGACCTATTGTGATTTATTTTACATTCCA 97004

Qy         209 ------------------------------------------------AsnLeuAsnLy 212
                                                               |||||||||||
Db       97005 GTCTCAGATAGCTAACCAATATGGGTTTCAACCATTCTGTGTGCTTTAGAACTTGAATAA 97064

Qy         212 sGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGluProTy 232
               ||||||||||||||||||:::    ||||||||||||||    |||    ||||||
Db       97065 GCAGGCATATGACCTGGCCAGGACTCTTCTAAAGAGAACGGTACAGACTATCGAAACCTG 97124

Qy         232 rIleThrThr-------------------------------------------------- 235
                |||    :::
Db       97125 CATCGCCTCAGTGAGTCACAAAATCATCTGTTTGTAATTTTATCTTTTGAGAATAAGTTT 97184

Qy         235 ------------------------------------------------------------ 235

Db       97185 CAGCATTATAGCATGTCAATATTATTTCATATGTTTGATTGGCTTGGCATTTCTTTTTCT 97244

Qy         236 ----PhePheAsnGlnValLeuMetLeuGlyLysThrSerIleSerAspLeuSerGluHi 254
                   |||||||||||||||||:::::::||||||:::|||:::||||||||||||||||||
Db       97245 GTAGTTCTTCAATCAGGTGTTGGTCATGGGAAAGTCCTCGGTGAGTGATCTGTCAGAGCA 97304

Qy         254 sValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuLeuLeuSerValLe 274
               |||||||||||||||||    ||||||:::    ||||||    |||||||:::|||||::
Db       97305 TGTATTTGACCTCATTCAGGAACTCTTTGCAATTGACCCTCTGCTCCTAGTCTCTGTGAT 97364

Qy         274 uProGlnLeuGluPheLysLeuLys----------------------------------- 282
               :||||||||||||||||||||||||
Db       97365 GCCACAGCTTGAGTTTAAATTGAAGGTGACCAAATTGACTCACACTTAATTTCACACCAG 97424

Qy         283 -------------------------------------------------------SerAs 284
                                                                      |||||
Db       97425 CTTGCATACAAGGCTCTTTAGAGTAAGTCTTTTTGTTTTGTTTTTGTCTTTGTAGAGTAA 97484

Qy         284 nAspAsnGluGluArgLeuGlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAs 304
               ||||    ||||||||||||    ||||||||||||||||||||||:::|||||||||||||||
Db       97485 TGATGGAGAGGAAAGGTTAGCTGTGGTAAAATTGCTAGCCAAACTCTTTGGCGCAAAAGA 97544

Qy         304 pSerGluLeuAlaSerGlnAsnLysProLeuTrpGlnCysTyrLeuGly----------- 320
               ||||||||||||:::||||||:::||||||||||||||||||:::||||||
Db       97545 CTCTGAACTGGCCACTCAGAATCGCCCACTCTGGCAGTGTTTCCTGGGAAGGTAGGATGG 97604

Qy         320 ------------------------------------------------------------ 320

Db       97605 CTTTTACCATTTAAAAATCTGTAATTAAAGTGTGAGAGATTTGTAGGATTTCTTAATGAA 97664

Qy         321 ------------------------------ArgPheAsnAspIleHisValProIleAr 330
                                             ||||||||||||||||||||||||:::||
Db       97665 TTTGGCTTTAATTTTTTGCTTTTTGTCTTGAAGGTTCAATGACATCCATGTTCCTGTTAG 97724

Qy         330 gLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaLysAs 350
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       97725 ACTGGAATGTGTGAAGTTTGCCAGTCACTGCCTGATGAACCATCCAGACCTTGCCAAGGA 97784

Qy         350 pLeu-------------------------------------------------------- 351
               ||||
Db       97785 TCTTACAGGTCTGTTCTGCTTGAATACAATGTCTGTTTGTTTTTTCCTATTTTAAGCACT 97844

Qy         352 ------------------------------------------------ThrGluTyrLe 355
                                                               ||||||:::||
Db       97845 CCTTGACATAGAGATGTCTTGTTTAACTTATGATGTGCTCTATTTCACTACAGAGTTTTT 97904
```

```
Qy        355 uLysValArgSerHisAspProGluGluAlaIleArgHisAspValIleValSerIleVa 375
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||:::
Db      97905 GAAGGTGAGATCACATGACCCAGAGGAGGCAATCAGACATGATGTCATTGTAACGATTAT 97964

Qy        375 lThrAlaAlaLysLysAspIleLeuLeuValAsnAspHisLeuLeuAsnPheValArgGl 395
              :   |||   |||||||||:::   ||||||||||||   ||||||   |||||||||||
Db      97965 CAATGCAGGGAAGAAAGACCTGAACTTGGTGAATGATCAACTGTTGGGTTTCGTCAGAGA 98024

Qy        395 uArgThrLeuAspLysArg----------------------------------------- 401
              ||||   |||||||||||
Db      98025 GAGGATGCTGGACAAGAGGGTAAGAAAGATCGTTGCCCTACAACAAAAAAAGATTTCAAA 98084

Qy        402 -----------------------------------TrpArgValArgLysGluAlaMetM 410
                                                 |||||||||||||||||||||||||
Db      98085 TCCACATATAACGTAAAATTTTGTTTCTTTTTAGTGGCGTGTCCGTAAAGAAGCAATGA 98144

Qy        410 etGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyLysAspAlaA 430
              |||||||||||||:::::::||||||||||   |||         ||||||||::::::|
Db      98145 TGGGTTTGGCTCAGCTGTTTAAAAAATACTGTCTGCACCACGAGGCTGGGAAGGAGTCTG 98204

Qy        430 laLysGlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrGlnAsnSerIleA 450
              ||   :::|||:::||||||||||||||||||||||||||||||||||||||||||||||
Db      98205 CTTTGAAAATCAGCTGGATCAAAGACAAACTCTTACACATCTACTACCAGAACAGCATAG 98264

Qy        450 spAspArg---------------------------------------------------- 452
              |||||
Db      98265 ATGACAA-GTAGGAGCCACACACATGTACACACTGATGTTCATGTACTTCTCTTACATAT 98323

Qy        452 ------------------------------------------------------------ 452

Db      98324 TCATTATCATAGTGTTGCATATCAAAAAAGGAAATCTCAGGTCCTTTAAGTGTTTCTTTA 98383

Qy        453 -----LeuLeuValGluArgIlePheAlaGlnTyrMetValProHisAsnLeuGluThrT 471
                   ||||||||||||:::|||||||||||||||||||||||||||||:::||||||||
Db      98384 TTAGGTTGCTGGTGGAGAAGATCTTTGCTCAGTATATGGTTCCTCACAGTCTGGAGACTG 98443

Qy        471 hrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspLeuAsnAlaValLysA 491
                 |||:::||||||||||||||||||||||||||||   ||||||   ||||||||||||
Db      98444 AGGAGAAGATGAAGTGTCTGTACTACCTTTATGCCTGTCTGGACACCAACGCAGTCAAGT 98503

Qy        491 la---------------------------------------------------------- 491

Db      98504 G-AGTCTCAATCCACTGCTTCATTTCCTTTTTAAAAGCTTCATCAGATTGTGAAGAGGC 98562

Qy        491 ------------------------------------------------------------ 491

Db      98563 AGTTTCAACAACACTCATCTCACTGATGTTTAGTTGTATTAAGCAGAGCATCTCTTTGT 98622

Qy        492 -----------LeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnValLysA 508
                         |||||||||||||||||||||||||||:::|||||||       |||::::
Db      98623 TTTTCAGAGCACTGAATGAAATGTGGAAGTGTCAGAATATGCTGAGAGGGCTGGTTCGAG 98682

Qy        508 spLeuLeuAspLeu---------------------------------------------- 512
              ::|||||||||||
Db      98683 AACTGCTGGACTTACACAAACTGCCTACAGTAAAGTCAAACAAGCATTCTTGAAGCTATT 98742

Qy        512 ------------------------------------------------------------ 512

Db      98743 TTTGGAACTTACACATCCACAAACACACTGAATTAAGGTCTTGATACCTTACTAAAGTTA 98802

Qy        512 ------------------------------------------------------------ 512

Db      98803 ATACTTGCCACAAATGTACGATTTACTCTTTCATAATTATTTTGTAAGTTTATGCTTTTG 98862

Qy        513 ----------------------IleLysGlnPro-------------------------- 516
                                    |||   ||||||
Db      98863 TCCTGACAAACAAAAAGGGAACTATCCA-CAACCTTTGCTGTATAATGACATTAAATATA 98921

Qy        516 ------------------------------------------------------------ 516

Db      98922 GTTCACAGTGGCTATGTCTACAAAAAGGTATTGTACATGTCTGTATGTCTTATGTGCAAT 98981

Qy        516 ------------------------------------------------------------ 516

Db      98982 ATACTATATCTTAGTCTTATTGTTGGTTTTTAATAAGAAAAAAATATTTTTGTCCTAAAT 99041

Qy        516 ------------------------------------------------------------ 516

Db      99042 ATAGTTTATTCAATATGTAGTTACTGAATATATATATATATATATATATATATATATATA 99101

Qy        516 ------------------------------------------------------------ 516

Db      99102 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 99161

Qy        516 ------------------------------------------------------------ 516

Db      99162 TGTGTGTGTATGTATGTATTTGCAGACATTAGTCCTTTCACGTGTATGATCCCTCCAGTG 99221

Qy        516 ------------------------------------------------------------ 516

Db      99222 TGATGAGCCTTGGCTGGTCTCTGATGCGTATGATCACACTGGTGTGATTAAAACATTCAC 99281

Qy        516 ------------------------------------------------------------ 516

Db      99282 AACAGAAAAGCTCAATTGAATTATTTCTGCAGTTTCAGTGATACTGACAAGAGTCGACAG 99341

Qy        516 ------------------------------------------------------------ 516

Db      99342 TAGCAGAGACTCGACTTAATGTGCTGTATGTGATCCACTTAAATTTGAGCCGCTTAATGT 99401

Qy        516 ------------------------------------------------------------ 516

Db      99402 TCCTGTGTAAACTCTGATTTAAACTGTTTAAAGCATACATCAATGCATGTTCTGCACTTA 99461

Qy        516 ------------------------------------------------------------ 516

Db      99462 AAAACAGACTGAATTTGGTACACGTGCACTATACCGAAAGTCCTGTACTGACCAGTTTAC 99521

Qy        516 ------------------------------------------------------------ 516

Db      99522 AAATATACCCAGAACATTTTACTTTCCTGTCTGTGGTTTACATGCTAAAGCATAAATACG 99581

Qy        516 ------------------------------------------------------------ 516

Db      99582 AATGACATACTGTAGCAGACAAAATATCTGCAAGTGATGAGACTATACTAACTAATTAGT 99641

Qy        516 ------------------------------------------------------------ 516

Db      99642 AGTGAAAAATGGTTGTGATGTTCCTATTTTGTAAGTGTGTACAATGAGCACAGGTTTTCC 99701

Qy        517 --------------------------------------LysThrAspAlaSerValLysA 524
                                                    :::::::::|||:::      |
Db      99702 ATATATGGATGACTGACTTGTTTTTCCCCTTTTCCGGCAGTCTGAAGCAAACACCTCAG 99761

Qy        524 laIlePheSerLysValMetValIle---------------------------------- 532
              ||:::|||   |||:::|||   |||
Db      99762 CCATGTTCGGGAAACTTATGACCATCGCTAGTAAGCAAATTGCATATATTTCTGTAAATT 99821

Qy        533 -----------------------------------------------ThrArgAsnLeuP 537
                                                             :::   |||||||
Db      99822 GATAAATCGACAGGTTTAACAATGATTAATTTGCCTCTCTCTCATGCTCAGAGAATCTAC 99881

Qy        537 roAspProGlyLysAlaGlnAspPheMetLysLysPheThrGlnValLeuGluAspAspG 557
              |||||||||||||||||||||||||||||||||||||   |||||||||   ::::||||
Db      99882 CAGACCCAGGCAAAGCTCAGGACTTTATGAAGAAGTTTAATCAGGTTCTGGGTGAAGACG 99941

Qy        557 luLysIleArgLysGlnLeuGluValLeuValSerProThrCysSerCysLysGlnAlaG 577
              |||||:::|||   |||||||||   |||:::||||||||||||||||||||||||||||
Db      99942 AGAAACTACGTCTTCAGCTTGAGCAACTCATCAGCCCCACCTGCTCATGTAAACAGGCCG 100001

Qy        577 luGlyCysVal------------------------------------------------- 580
              ||    ||||||
Db     100002 AACAGTGTGTGGTAAGCTTAAAGAGCCCATATTATGGGTTTTGAAAATTCCCCTCCATG 100061

Qy        580 ------------------------------------------------------------ 580
```

```
Db     100062 TAGTGTGTAACACAGCTCTAAGTGAAGTGAAGTATCCAGCTAAGGCTTAAATCTGTAAGA 100121
Qy        580 ------------------------------------------------------------ 580
Db     100122 ATACAGTGTTTAAAACTGTTGATTCATCTATAAAAGAGTCGAATCATAGTGCTTCAAACG 100181
Qy        580 ------------------------------------------------------------ 580
Db     100182 AGTCGCCTTGATACCGACTCATTAGGTGTTTCGCCATGACGTACGAACGAAACAAAGTTT 100241
Qy        580 ------------------------------------------------------------ 580
Db     100242 TTCACGTGCACGCGCAAACCCGGGAGATTTCAAACCTGAGGCCCCGCCCTCTGACGCAGT 100301
Qy        580 ------------------------------------------------------------ 580
Db     100302 AACCCAGACACACACACACACACACCCACAAACACACGCACACCCACAAACACACGCACA 100361
Qy        580 ------------------------------------------------------------ 580
Db     100362 CAAACATGCCGGTCGATTGAAGTCACACTGCAGATGGATATATTGAGTCTCTATCCAAAG 100421
Qy        580 ------------------------------------------------------------ 580
Db     100422 ATGAAACCTCAGCATTATAACCAAGCAGTTGGAAACTTCTGGAAGCTACATGCTACAAAG 100481
Qy        580 ------------------------------------------------------------ 580
Db     100482 AATACTTCATCTGAGTTTGTTAAAGGAAGGATCAGTAAAGAGTAACTGATGGACGTCAGG 100541
Qy        580 ------------------------------------------------------------ 580
Db     100542 ATGGGTTTCTTCCTACATTTCTCAAGTGTAAGTACGTGCGGTTAAAGTTGTTGCCTCGTT 100601
Qy        580 ------------------------------------------------------------ 580
Db     100602 TACTCTAGCTTGCAAATGTATTTAGTTGTGATTTGTTACTTGTAACCGCGTGTACTGTAT 100661
Qy        580 ------------------------------------------------------------ 580
Db     100662 CAGGTTAACTGGATATTTGATCTTATCGCGTGCAAAGTCACGTTAAAAACGCGACGCGTG 100721
Qy        580 ------------------------------------------------------------ 580
Db     100722 CTGTTTGTTTATGGAGCTCCGTGCTAGAGTTCTGCTCCCGCGATTTATTCGGAAATGTAT 100781
Qy        580 ------------------------------------------------------------ 580
Db     100782 TCCCGCGCCGCAGAAATCTGGCGCGGGGAGAGAGAGAGAGAGAGAGAGAGCGGTAGGGAG 100841
Qy        580 ------------------------------------------------------------ 580
Db     100842 AGCGAACGAACGAACGAGCTTTGTGTACTTTGTGCTGTGTGTGTGTTTTTATACTGACAG 100901
Qy        580 ------------------------------------------------------------ 580
Db     100902 CTTGGCTGTCTGGACTGTATACTGGGTGATATTTGGTTGTGTTCTCCGCTCTAGCGGGAC 100961
Qy        580 ------------------------------------------------------------ 580
Db     100962 CGGGCGCGGCGGGCAGAAAACGGAGCGGGTTCTCAGGCTAAAACCTGGCGGGTGCGGGCT 101021
Qy        580 ------------------------------------------------------------ 580
Db     101022 GAAGCGGGAGCGTTGTCATCCGGAGGTGTGTGTGTGTTTTTGTGTGTGTGTGGTATGGCG 101081
Qy        580 ------------------------------------------------------------ 580
Db     101082 AGTACACGACTGTCTCCTTCGTTCGGTTATCCGTGGCACTATCCACTCGCCATAGTGTGG 101141
Qy        580 ------------------------------------------------------------ 580
Db     101142 CAGCTAAAGTCCACGCCTACACTATCGAGCGTGTATGAACTGTGATTACTTTTTAAATTG 101201
Qy        580 ------------------------------------------------------------ 580
Db     101202 CTGATTAGCTATTGGCCATTTCCCTCTCTGACTGAAGGCAGTCGACCAATCGCGACAGAC 101261
Qy        580 ------------------------------------------------------------ 580
Db     101262 TGTCATCGGTCCAATCAGCGCAGATTAGCTTCGCGCTAAGGAGGGGTTTGGGAACAAATG 101321
Qy        580 ------------------------------------------------------------ 580
Db     101322 AATCACTGGACGATTCATACAGGAGTCGCTGGGATAATTAGATAAAAATAAATGCAGATT 101381
Qy        580 ------------------------------------------------------------ 580
Db     101382 ATAAGACCATGAAAGTGTTTTTTGACCTTGCATGCATATTAGACTGTTGTTGGAGACCCT 101441
Qy        580 ------------------------------------------------------------ 580
Db     101442 TACAACCAAGATATGACCCTATTTCATGTATAATATGGGCTCTTTAACAAAACTTTTTTT 101501
Qy        580 ------------------------------------------------------------ 580
Db     101502 AATGCTTGTTCAAAACTCCGGCCATGCTTGTCATAACCAGGGCTTCACATGAACTTTTTT 101561
Qy        580 ------------------------------------------------------------ 580
Db     101562 GATCACCAGCCAACAGTGGCTATTAGTTTTTCAACGTTACTAGCCACTCTGCATTTCACT 101621
Qy        580 ------------------------------------------------------------ 580
Db     101622 AGCCACAATTTTGTTGTTGGGAAAATATTTTTTATATGCACAAATATTACTTTGACATGC 101681
Qy        580 ------------------------------------------------------------ 580
Db     101682 TAAAATTACTTGATTTAGATTTTGTGTTTTGTTCACATGCCGCCTCATTAATTGCACGTT 101741
Qy        580 ------------------------------------------------------------ 580
Db     101742 TTGTGTGTCGTGTATGAGCAAATGAGTCATGGTTTCATAGTGTTCTATTACAATTAGTTT 101801
Qy        580 ------------------------------------------------------------ 580
Db     101802 TTATTTCACAGGATTTTTCAGAACTCAAAATTAGGGATGTACCAAATTTATCTCTGACCA 101861
Qy        580 ------------------------------------------------------------ 580
Db     101862 CTCCGAACAATAATTATTTCTTAGCCACACGTTTTAAATGCGTCAAAACAAGCTAAATAT 101921
Qy        580 ------------------------------------------------------------ 580
Db     101922 AGCTGTATATTATACTTTTTATTTAACAAAACATTTCTTAAAAAAAAGAAAAAAAAAAAG 101981
Qy        580 ------------------------------------------------------------ 580
Db     101982 AAAAACCATTGACTTTTAAAAATAGTTATAGTCATGTATCAAAAATATGCAGAGTAATCT 102041
Qy        580 ------------------------------------------------------------ 580
Db     102042 GTCCTCGCTAAAGGTCTCTCAGATCACTAGTTACTACACAAGTGGAGAGTTAATCTCATA 102101
Qy        580 ------------------------------------------------------------ 580
Db     102102 TTAAAGCAATGTTGTTGTAAAAAAATGATGGATAAACATATAAACAAAAATAACCGTAAG 102161
Qy        580 ------------------------------------------------------------ 580
Db     102162 CAAAAGAAAGCAGGTGAAAAACATACTGTGTGATAATGAAAGGATGATCGCATGGTTTTG 102221
Qy        580 ------------------------------------------------------------ 580
Db     102222 GCCAATTAATAATCTGTTCAAAGTGAAAGCAACTGTTGCTGCTTACAAAAAGACCAATCT 102281
```

```
Qy        580 ------------------------------------------------------------ 580
Db     102282 GGAGCTCTTGAAAGCAGCAAGACTGCGGGTGCATTAGCATAACTTTTTGTCTGTTCTATA 102341
Qy        580 ------------------------------------------------------------ 580
Db     102342 TGAAAGAGAACTGATAACAGTTGCGTTCCAGGCACGATTGCTTCAAGCAAACAAACAGGA 102401
Qy        580 ------------------------------------------------------------ 580
Db     102402 GCGCACACGCGGTTTATAAAGTCGCGTGTATCATCAGCTGTGTGCACGACTGATCATCCT 102461
Qy        580 ------------------------------------------------------------ 580
Db     102462 CTCATTCATTCAGTCCCAGATCATCCAGGTATTTACCTGCTTTCTATTGCTCGCGCGAAA 102521
Qy        580 ------------------------------------------------------------ 580
Db     102522 GCAAATTATATTTGTTAGTCGTGCTGTTTCTCGATTCTAAGATCACATTTGCACGCATAT 102581
Qy        580 ------------------------------------------------------------ 580
Db     102582 TGGGTCATCCTTATAGTCGAACTCTGCTTTTAAGAAAACTACAATTTAAAAATTGTTTTC 102641
Qy        580 ------------------------------------------------------------ 580
Db     102642 AACCAGCTGAAGTGGCTAGTGGGTAGGGCTGCACGATATATCGTTTCAGCATCGAAATCG 102701
Qy        580 ------------------------------------------------------------ 580
Db     102702 CGATGTGCGCATCCGCGATAGTCACATCGCAGGAGGGTGCGATGCAAAGCAAAAAAAAAA 102761
Qy        580 ------------------------------------------------------------ 580
Db     102762 AAAAATTCTATATATATTATGGGTGGAGCCGAACCGGAATATGGTATTCAGAAAAGCACG 102821
Qy        580 ------------------------------------------------------------ 580
Db     102822 AATATATGTTTTTACGAATACTTATTTCAAACAATACTTGATAAATTATTTGTATTTGGG 102881
Qy        580 ------------------------------------------------------------ 580
Db     102882 AACAAGAAAAACATTATATCAAAAAGCTGCGTTTCCTAATGAGACCGCAGTGCATGCCCG 102941
Qy        580 ------------------------------------------------------------ 580
Db     102942 CGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGATGTTGGGGTGGGGCCGGGGTA 103001
Qy        580 ------------------------------------------------------------ 580
Db     103002 AGGATACTGAGCAAGCTGCTTTTCTATACAGCAACAGAGAAGGCTTGGCAGAGCGAAAAA 103061
Qy        580 ------------------------------------------------------------ 580
Db     103062 TACTTTTCTGCATTACTATTTTTGTTTATTTTATCATACTAAAATAACACGTGACAGAAG 103121
Qy        580 ------------------------------------------------------------ 580
Db     103122 TTCACTCAAAAGTAAAAGTTCTTTCTGTTATTAATTACTAACACTCCCGTCATTCCTACT 103181
Qy        580 ------------------------------------------------------------ 580
Db     103182 ACTTTCAGAACATAAATTTTGATATTTTACCATATGAAAATGCAATGCAGTTATGAAAAT 103241
Qy        580 ------------------------------------------------------------ 580
Db     103242 TGAAAAATTAAAGCAATACACAGATCTGGTCCCCGCGTCAGGTCAGGGTGCACCTGTAAC 103301
Qy        580 ------------------------------------------------------------ 580
Db     103302 CTTAAAAACGCAGACGCGGCGGAGGCGCACTAAACTCACACGGGAAAGAACTGGGCATTC 103361
Qy        580 ------------------------------------------------------------ 580

Db     103362 AATTCATGTTTTCGTTAAATAATTCTTTAAATACAGCTCTTTTGATCTCTAAAGGGAAAA 103421
Qy        580 ------------------------------------------------------------ 580
Db     103422 AAAACGCAAAAGAAATTTGTCGTTAAATTAATTTTAGCAAAGACGCAACTGCCTGTCATT 103481
Qy        580 ------------------------------------------------------------ 580
Db     103482 TTCACATTCGTTTTTCAGTTAGGGATTTGGTCTATTAGAGTAACCTAATAAAATTGTTTT 103541
Qy        580 ------------------------------------------------------------ 580
Db     103542 GTTGGACACTGTAGTCTACACAATAATAAATATTATAATTTATTATATACATTATTATTA 103601
Qy        580 ------------------------------------------------------------ 580
Db     103602 AAACTAAACTATACAATTGATAAAACAAAACCCTTTTTAATGTTTGCTTTCATTTTGACA 103661
Qy        580 ------------------------------------------------------------ 580
Db     103662 CCGCTAACAGTTGTAATAAGAAACATCACATTATGCAGAAGCATGCAGTCTCAAGTTAAA 103721
Qy        580 ------------------------------------------------------------ 580
Db     103722 AACACCTTTTGAGTTTAAAGATGATCAGCCTGCAAAGTCAAGTTTACAGTGTCATAAATG 103781
Qy        580 ------------------------------------------------------------ 580
Db     103782 AAATGGGTAGCATGAGTAAAGGGGATTTGCCCAGCACACAACCTCATGCGCTGCAGGTAA 103841
Qy        580 ------------------------------------------------------------ 580
Db     103842 GCTGAATGATAATCTTAAAAAATATATAGGCTATGCAATACAAAACTTTAAACATTTATT 103901
Qy        580 ------------------------------------------------------------ 580
Db     103902 TATTTTTTAATGAAAAGAAAAGAAAACAGGAGAGTAAAATATAATTTAAACCACTGGTCT 103961
Qy        580 ------------------------------------------------------------ 580
Db     103962 GGCGCTGATACTTTATTGTTTGTATCCAGAAGAAGGGGGAAAAAGCAGATGAAAACCTGG 104021
Qy        580 ------------------------------------------------------------ 580
Db     104022 CTGGGGAGACGGCGATAACATATCAGTTCTTCAGCAAAGATCACGACATCATTGTGATAC 104081
Qy        580 ------------------------------------------------------------ 580
Db     104082 GTTTATCATACAACGCAGTTCGTGTTTATAATCATTAATATTAACTAATATTAATTCTCT 104141
Qy        580 ------------------------------------------------------------ 580
Db     104142 CTCTAGACTTAGTTTTAGGGCTTGTTCTTGTCGGCGCAATATAGCCTCGGCTTATTGCAC 104201
Qy        580 ------------------------------------------------------------ 580
Db     104202 GTACGAATCCGATTGTGAAGAATTATTAAGCGCTGGCATTTTTATTGCATCCTTTTCGTA 104261
Qy        580 ------------------------------------------------------------ 580
Db     104262 AGTTATTCATTTTTTTTTAAGTGGTGTGAACATTTTTTACTGGCAAACATGTTCCATGAC 104321
Qy        580 ------------------------------------------------------------ 580
Db     104322 CTCGTACATCAACAAATAATGTAGCCTACTTAAGTTAATTTACCTTTTATGTAAAAATTA 104381
Qy        580 ------------------------------------------------------------ 580
Db     104382 AGGAATAACTTAATAAAATATTAATAAATAAAACATCCGTATAAAGTACATGTTTATTTT 104441
Qy        580 ------------------------------------------------------------ 580
```

```
Db     104442 TGAACTTTCGTTAATAAAACAACACAAAACTGTTTATAAATATTGTAGTTAAATATTAAA 104501
Qy        580 ------------------------------------------------------------ 580
Db     104502 TTACTTTAACTGCTTGAAACCGTATAATAATAACAATTAAACGTTACATTTTTAAGAACT 104561
Qy        580 ------------------------------------------------------------ 580
Db     104562 TAACAAATAGTCTAAATGGCACTTTACTAATCAGTGTAATAATAATATAAATATTAATAA 104621
Qy        580 ------------------------------------------------------------ 580
Db     104622 TTATTAGTAAAAAGAAAATATTAACCTAGTAACAATATAAGTCAGATGTTTAAAAAAAAC 104681
Qy        580 ------------------------------------------------------------ 580
Db     104682 TGCGTAGTTTATCTCCACCTTTTGTATTCATATTTAAATTATTTGCCAGAAAAACTAAC 104741
Qy        580 ------------------------------------------------------------ 580
Db     104742 ATGTTGACTTTGTCCGGTTTAAGTGAAGAATGGGGTTATAATGTTTCCAGCAATACACTT 104801
Qy        580 ------------------------------------------------------------ 580
Db     104802 TCAGACGGTGTGCTTGTAGCGAAAATGCACAGACAGGTTACTTTTTCACAACCATGCGAG 104861
Qy        580 ------------------------------------------------------------ 580
Db     104862 AGAGGGCTTCTGCTTCTGTTCCGCCGCCCGCATTTCTTGCAACGCAGGTTTCTCTACGCT 104921
Qy        580 ------------------------------------------------------------ 580
Db     104922 GCGGAGCGGCGCGTTCTTAAGTAGTTGTTACCAAATAGCCTATGTTGCTCCATTTCTCGT 104981
Qy        580 ------------------------------------------------------------ 580
Db     104982 CAATGTCTTCCATTTATCACTTTCTCTTTTGTATCTGGCCATTTTTGCAAATGCCTCTGC 105041
Qy        580 ------------------------------------------------------------ 580
Db     105042 CATGCCTGGCTGAGCTGGGCTTCTGCGTTTAGAAGACAGGAGTTGATTAAGCCTACATGG 105101
Qy        580 ------------------------------------------------------------ 580
Db     105102 CTTTTTTTCCCTCGCCGTCAATCGCCACTAGCTTGCGTGGTTCGGGACCCGTGGATCTGC 105161
Qy        580 ------------------------------------------------------------ 580
Db     105162 GTATAGATGGATTTGCTCCTCAGTGTTTGGACTCTCAGCAGTGAATATTAAAACACACTT 105221
Qy        580 ------------------------------------------------------------ 580
Db     105222 TACTTAAAACCCCCATACAAAAAGCTTAATCCAAATTGATTCGTTCACCTTTCTCAGTCG 105281
Qy        580 ------------------------------------------------------------ 580
Db     105282 GCACATATCCAAAATGTTCCCAGATGTAAGAGTTCGTTTTCTAACAAAATTCATGGTTGA 105341
Qy        580 ------------------------------------------------------------ 580
Db     105342 ACTTGAACTTGCTGAATCGCGCCTAGTGTTATATTCTAACAGATATGTAATACTTGGCAA 105401
Qy        580 ------------------------------------------------------------ 580
Db     105402 CATTACAGTAATTTGTGTTTTAAAGTGCGTGACGTCACGTACTACGGCCGCTGGCAATGG 105461
Qy        580 ------------------------------------------------------------ 580
Db     105462 ACAACCGAGGTGGCAACTGACCACCGCGGTGACGCGGTTGTCATGGTGACCGTCACAGCC 105521
Qy        580 ------------------------------------------------------------ 580
Db     105522 CTAGTTTTAAGTATATATATATATATATATATATATATATATATATATATATATATATAT 105581

Qy        580 ------------------------------------------------------------ 580
Db     105582 ATATATAATATCGCAATATATATCGCAGGGGAAAAAAATATCGCAATGTCAATTTTTTCC 105641
Qy        580 ------------------------------------------------------------ 580
Db     105642 AATATCGTGCAGCCCTACTAGTGGGAGTGACTGTCTAACCTGCCAAAGCTGAAATCTACC 105701
Qy        580 ------------------------------------------------------------ 580
Db     105702 CACATTTGGCGGGTTGGCGGGTGTTAATGTAAAGCCCTGGTCATAACCATTCACAATCGT 105761
Qy        580 ------------------------------------------------------------ 580
Db     105762 CCTGTTTAAAATGATTACAATTGACCTATGTTCTGATCAACAATTAAGCAAGCTTTATTA 105821
Qy        581 ------------------------------------------------------ArgGlu 582
                                                                    ||||||
Db     105822 TAGTACTAAATATTAAACAGACATAAGTTATTTGAAAGTGGCATGTCGTTTCAGAGGGAA 105881
Qy        583 IleThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLys 602
              ||||||:::||||||      |||||||||||||||||||||||||||||||||:::|||
Db     105882 ATCACACGAAAGCTGACATTCCCTAAGCAGCCGACAAATCCATTCCTAGAGATGGTGAAA 105941
Qy        603 PheLeuLeuGluArgIleAlaProValHisIleAspThrGluSerIle------------ 618
              ||||||||||||||||||||||||||||||||||||:::|||:::|||
Db     105942 TTCCTGCTGGAGAGAATTGCACCAGTGCACATTGACTCTGAGGCAATAAGGTAAATGAGA 106001
Qy        618 ------------------------------------------------------------ 618
Db     106002 ACTGTTTTTTTTTTCTGTTGGAAAGAAAAGTTTCATATAGCACTCATGCCGCAGATGAAT 106061
Qy        618 ------------------------------------------------------------ 618
Db     106062 AAAGTATTGTCTTAAGTCTTTTTTTATTATTAGTTTAATAAGATTGGCAGCAACAAAAAA 106121
Qy        618 ------------------------------------------------------------ 618
Db     106122 AAAAAGATTGAAAATTTTATTTTCATTATAAATGAAAGTATTATCTGATGTTTTAATTCT 106181
Qy        618 ------------------------------------------------------------ 618
Db     106182 TTCAAGAATACTAATAAAAAAGTCTTTGCTATAACAGCAATCATAAATGTACTGTACATG 106241
Qy        618 ------------------------------------------------------------ 618
Db     106242 GACAAGATTAAGATAACTGACAACACATTTGCTCATGAACTCCTACTTTCAGTTATACAA 106301
Qy        618 ------------------------------------------------------------ 618
Db     106302 AACAACATGCAAGTACCTTATGTGTGAACTAACAAATTTGTTCAGTTACAAAATAAGTGT 106361
Qy        618 ------------------------------------------------------------ 618
Db     106362 TCAGTCAGTTTTAAATCAAATTGTAAATATATACAGAGTGCTGTTAAATGGCAGTTCTAT 106421
Qy        618 ------------------------------------------------------------ 618
Db     106422 GGAGTCCTATCAGTGCACCTGCAGAAGTGAGGTGAGAGCATGAATGAGAGAGGAAGAGAA 106481
Qy        618 ------------------------------------------------------------ 618
Db     106482 ACTGACTGAGTGGGAGAAAGATAAGCACTATATGGGCCATGTGCTTTTTCCCTCCCTAAC 106541
Qy        618 ------------------------------------------------------------ 618
Db     106542 GGTTGTGTAACAGCCTCAATGGCTTTTTAACGATGTGACATCTAGTGAGGAAAAAGTGTA 106601
Qy        619 -------------------------------------------------SerAlaLeuIl 622
                                                               |||||||||::
Db     106602 ACTGCAGGTATTGTGTGGCCTTTTTGATAAACACATTATGTGTGTCTACAGTGCCCTGGT 106661
```

```
Qy        622 eLysGlnValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProTh 642
              :|||   :::||||||||||||:::|||||||||||||||||||||||||||||||||||
Db     106662 TAAGCTGTTAAATAAGTCAATTGAAGGCACTGCTGATGATGAAGATGAGGGAGTAACTCC 106721

Qy        642 rAspGlnAlaIleArgAlaGlyLeuGluLeuLeuLys----------------------- 654
               |||   ||||||||||||||||||||||||||||||
Db     106722 TGATACTGCCATCCGTGCTGGACTTGAGCTACTTAAGGTAATGTTTGCAATTGCAAATGT 106781

Qy        654 ------------------------------------------------------------ 654

Db     106782 AGTAGTTTTTATTTTAAATAAGACTTTTAAATACCATGTGCGCATTTTTTTCTTTCCA 106841

Qy        655 --------ValLeuSerPheThrHisProIleSerPheHisSerAlaGluThrPheGluS 672
                      ||||||||||||||||||||    :::||||||||||||||||||:::||||
Db     106842 TGCCTCAGGTTTTATCATTTACTCACCCCACTGCGTTCCACTCGGCTGAAACATACGAGT 106901

Qy        672 erLeuLeuAlaCysLeuLysMetAspAspGluLysValAlaGluAlaAlaLeuGlnIleP 692
              ||||||||   ||||||||||:::|||:::||||||||||||||||||||:::|||||||
Db     106902 CTCTGCTGCAGTGTCTGAAAATGGAGGACGATAAAGTGGCTGAAGCAGCAATACAGATCT 106961

Qy        692 heLysAsnThrGlySerLysIleGluGluAspPheProHisIleArgSer---------- 708
              ||:::||||||||||    ||||||||    :::    |||    |||||||||
Db     106962 TCAGGAACACAGGACAAAAAATTGAGACAGAACTGCCACAAATTCGCTCGTATGAACACC 107021

Qy        708 ------------------------------------------------------------ 708

Db     107022 TCTCATATTCATTTACATAAAAACAGCTCTACCAGAGTCAGAACTAGAAGCCTAAACTTA 107081

Qy        708 ------------------------------------------------------------ 708

Db     107082 TTAAAACCCTGATACATGCTTATTAAAATACAAATATAATGTAGTGAAATGTTTTTTTGT 107141

Qy        709 ---------------AlaLeuLeuProValLeuHisHisLysSerLysLysGlyProPro 723
                             |||:::|||:::||||||    |||:::|||:::|||    |||
Db     107142 AACTTTATGTGCAGGACTCTAATTCCCATACTTCATCAGAAGGCAAAAAGAGGAACCCCG 107201

Qy        724 ArgGlnAlaLysTyrAlaIleHisCysIleHisAlaIlePheSerSerLysGluThrGln 743
                 |||||||||    |||:::|||||||||||||||||||||    :::||||||    |||
Db     107202 CACCAGGCTAAACAGGCGGTCCATTGCATCCATGCCATCTTTCACAATAAAGAGGTCCAG 107261

Qy        744 PheAlaGlnIlePheGluPro--------------------------------------- 750
                 |||||||||||||||
Db     107262 TTAGCTCAAATATTCGAGGTAAACATCTTTTTTTGAATGAAAATATCACTGTTATCACAT 107321

Qy        750 ------------------------------------------------------------ 750

Db     107322 TTCAATTCGCATATACAGTAAACATTATAGAAAATGAAAGGGTCATGATTCCAACTATTT 107381

Qy        751 ------------------------------LeuHisLysSer------------------ 754
                                            ||||||||||||
Db     107382 CACCACTGAAAATTTTCAGTCATTTTGTTTTGCCTTCACAAATCAGTAAACTCTTGAACC 107441

Qy        754 ------------------------------------------------------------ 754

Db     107442 ACTATTATTAACAACTGTGATATTTTTGCTTTGGCAATTATTTACTTGTGGCGTGCTGTG 107501

Qy        754 ------------------------------------------------------------ 754

Db     107502 AGCAATTGACTAAACCATATTTCAAAATAATAGTATAATCTATTCAGAGAAAAAAAATG 107561

Qy        754 ------------------------------------------------------------ 754

Db     107562 TAATGAATGGTCCTTTATTAAGGAGGAAGGTCAATGCTGTGTTTCACATTTTTATAAAAT 107621

Qy        754 ------------------------------------------------------------ 754

Db     107622 ATTGTTATATGTATTACCTTCTGAAATGTGTTGTAAAATGAGTTGTTCCAAACACTGTTC 107681

Qy        754 ------------------------------------------------------------ 754

Db     107682 AAAGGAACACTAATTTATATTAAATGTTTATTTTAGAAGGGAGAATGTAAAAGAAATGCA 107741

Qy        754 ------------------------------------------------------------ 754
```

```
Db     107742 GCATATTGCATTGCAGGATGATCATCTGAAATGATCTAAAATGCTTTGAACTAGCAACTA 107801

Qy        755 ------------------------------------------LeuAspProSerAsn-Le 760
                                                        :::||||||:::|||  ::
Db     107802 AAACCCAAAACTTGTGAAAGAAATCAATCAACTACAGGTAAAATGGATCCAGCCAATTAT 107861

Qy        760 u----------------------------------------------------------- 760
              :
Db     107862 CACCTCTAGAAAGATCTTTATTTTTTATTTTTAGATCATCTTTTTCATCATGATTGTTTC 107921

Qy        760 ------------------------------------------------------------ 760

Db     107922 TACAACAACAGCTTTGTTTCAGGCATGAAGGATTAAGGTAATGGAGTGGCCAGCTCAATC 107981

Qy        760 ------------------------------------------------------------ 760

Db     107982 GCCTTACTTCAACCTAATTGAAACTCATTGGCTAAAATTAGGATTGTGCCGATAGACGAT 108041

Qy        760 ------------------------------------------------------------ 760

Db     108042 TGTATCAACGATAGGCAAAAATATCGCAGAGAGCTGAAACCTATGATGATTTTAAAGACG 108101

Qy        760 ------------------------------------------------------------ 760

Db     108102 ATATTTAGCTCCTTTCGCACTGTTACATGTTTTAATTGCATTATAAAATAAAATGAATGG 108161

Qy        760 ------------------------------------------------------------ 760

Db     108162 TATTACTATTTTTAATCATCATCATAAATAAAAAATTAACTATAAATAATATGATTGCTT 108221

Qy        760 ------------------------------------------------------------ 760

Db     108222 CAGCTGTTTATATCAACATCACCTGACAGACACTTTTGGACAAAAATGAATTATTTCAAT 108281

Qy        760 ------------------------------------------------------------ 760

Db     108282 TGCTCTTCTATCAAAAAACTTTTTACTGCCATCAGAAAAGTTTTAGGTCAAATAGGCTAT 108341

Qy        760 ------------------------------------------------------------ 760

Db     108342 TACTGACTCGCCGCAGCCATGTCTAAAATGAAACATATCATCAATATGGCGATCTGAAAA 108401

Qy        760 ------------------------------------------------------------ 760

Db     108402 TTTTGTATATTGCCCAACACTAGCTGAAATACCTGTTTCTAGGTGCCAGAAGTTATATTA 108461

Qy        760 ------------------------------------------------------------ 760

Db     108462 AAAGCAATATTAAAATATTCAGATAAATATTAATATCAATATTTTTTTCTCTAAACTGAA 108521

Qy        760 ------------------------------------------------------------ 760

Db     108522 ACACTAATAATAAATTAATTAATGCACCTTTTCTCTATCTTGTTTGATGGACAAGACAGT 108581

Qy        760 ------------------------------------------------------------ 760

Db     108582 GTCTGCTTGTTCATAAATGAAATGAAAATAGTCGACTGTACTTTAGCTGCATTATTTTGC 108641

Qy        760 ------------------------------------------------------------ 760

Db     108642 ACAGTAGATTGCTCCCTGTCACCTGATGGTGGAAAGCAGTGACTTGAGTGCTTGACATCT 108701

Qy        760 ------------------------------------------------------------ 760

Db     108702 ATTATTGATCAATCTAAAATCTTCATAGGATGGCAGACTGTTCTCTATATCAGCACACAA 108761

Qy        760 ------------------------------------------------------------ 760

Db     108762 AAGCCACATGTGGTGAAAGTTTTTAGAAAATGAAAATGGGTTGTACTACAATTGTTTGTG 108821

Qy        760 ------------------------------------------------------------ 760
```

```
Db   108822 CTCGCACAATTTCTACCAAATGTATTTTATGATTGCAAAGTTTACATTTCTGGTACATAT 108881
Qy      760 ------------------------------------------------------------ 760
Db   108882 GAGACTAAAATGGTTGTAATTTACAACCCTGACTATTTTTGTATCTTGTTGTAGGATAAG 108941
Qy      760 ------------------------------------------------------------ 760
Db   108942 AACATGCAGACTTTATTTATGAAACATCAGTGAGTCATCAATTAAAGTGGTCTTTTGCCA 109001
Qy      760 ------------------------------------------------------------ 760
Db   109002 CATCATATCCTGAGGTGTTGACGAGATTTTAAAACTAGATATTGCAAAAAAATAAAAAATA 109061
Qy      760 ------------------------------------------------------------ 760
Db   109062 AAAACAAAAGCACAATTAACCCACTTTCTCCTGCAAATATCAGCAGATTTATATAAAATG 109121
Qy      760 ------------------------------------------------------------ 760
Db   109122 GGTGCTAACTTTGCCTTTTTATGATGAAAAATGTGAACAAATGTATTTTAAATTGTTTAA 109181
Qy      760 ------------------------------------------------------------ 760
Db   109182 TTTTATTTTTATTTTTTTTATGCATTTGTTTTCTGCACGAGTTTTTTCAAAACATTTTTA 109241
Qy      761 ---------------------------------------------------GluHisLe 763
                                                                |||   ||
Db   109242 TCTCTGTGTTTTGTTCCAGCCGTTGTCTCGGAGTCTGAATGCAGATGTACCTGAACAGCT 109301
Qy      763 uIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGlnPheAlaAl 783
            ||||||||||||||||:::::||||||||::::::||||||||||||||||||||||::
Db   109302 GATCACTCCACTCGTGTCTCTGGGACACATCTCCATGCTGGCCCCGGATCAGTTTGCATC 109361
Qy      783 aProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsnAspArg-- 802
            :|||  ||||||  ||||||  ||||||||||||||||||||||||||||||
Db   109362 TCCCATGAAGTCAATTGTAGCTAATTTATTGTAAAGGACCTTCTGATGAATGACAGAGT 109421
Qy      802 ------------------------------------------------------------ 802
Db   109422 GAGCATATTGATGTATTATTCTGCTTTGAGCATGTGGAAATGTTTTTTATAGCCTTGCAT 109481
Qy      803 ---------------------------LeuProGlyLysLysThrThrLysLeuTrpVa 813
                                        |||  |||  |||   :::||||||
Db   109482 GTTAAATTTACCCTCCTCATCTCTTAGTCTGTG-GGAAACAAAAATGGGCGGCTCTGGAC 109540
Qy      813 lProAspGluGluValSerProGluThrMetValLys---------------------- 825
            |||:::||||||||||||||||  :::   |||
Db   109541 AGCTGATGATGAGGTTTCTCCTGAAGTTTTGGCTAAGGTAAAGTAATAATAATCATAATA 109600
Qy      825 ------------------------------------------------------------ 825
Db   109601 AAATTTCACATTGTGCCTTTTAAAAATTCCCAAAGCACTTTGCACTGAGAAAATACAGTT 109660
Qy      825 ------------------------------------------------------------ 825
Db   109661 AATACAATAATAAGGTAGAAAACGTCATTTTTGTATGTGTATTTATTTATTTGTTTGTGT 109720
Qy      826 --------IleGlnAlaIleLysMetMetValArgTrpLeuLeuGlyMetLysAsnAsnH 843
                    :::||||||||||||:::::::|||||||||||||||||||||||||||
Db   109721 GTTTCCAGGTGCAAGCCATTAAGCTGTTGGTCCGCTGGTTGTTAGGAATGAAGAATAATC 109780
Qy      843 isSerLysSerGlyThrSerThrLeuArgLeuLeuThrThrIleLeuHisSerAspGlyA 863
               |||||||||      ||||||||||||||||:::   :::|||   ||:::||||
Db   109781 AATCCAAATCAGCTAACTCTACTCTCAGACTGCTCTCAGCCATGCTGGTCAGTGAAGGAG 109840
Qy      863 spLeuThrGluGlnGlyLysIle------------------------------------- 870
            ||||||||||||||   ||||||
Db   109841 ACCTCACGGAGCAGAAGAAGATTAGGTACTCACTAGTGATCTAGAGTGATTTAGTATGAT 109900
Qy      871 ---------------------------------------------SerLysProAsp 874
                                                          ||||||   |||
Db   109901 AATAAAAGACAATATGCCTCTAACACACATTTGTATGTTTGTGTACACAGCAAGTCTGAC 109960
Qy      875 MetSerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCys 894
            ||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db   109961 ATGTCTCGACTAAGGCTTGCAGCCGGGAGTGCCATCCTGAAACTAGCACAAGAGCCCTGT 110020
Qy      895 TyrHisGluIleIleThrLeuGluGlnTyrGlnLeuCysAlaLeuAlaIleAsn------ 912
            ||||||:::|||||||||   ||||||:::|||||||||  |||   ||||||
Db   110021 TACCATGACATCATTACTCCAGAACAGTTCCAGCTCTGTGGACTCGTCATTAATGTGGGT 110080
Qy      912 ------------------------------------------------------------ 912
Db   110081 TCAGCTCTACTGGGGCATATATCACTTCAAATATTTAACTGAAAACAGGATAGCCTTGAC 110140
Qy      912 ------------------------------------------------------------ 912
Db   110141 TTTTAGGCAGCAAAAAACAAAATTTCCCTTTGCCACATGATGCAGTCTAATTCCAGGGTC 110200
Qy      912 ------------------------------------------------------------ 912
Db   110201 TGTGCACACATCAAATAGACAGGCACATTTTTGTTCAGATGTTTTGATGAAATGGCTGAA 110260
Qy      912 ------------------------------------------------------------ 912
Db   110261 GATCAATTTTAAAAGTAGACTTTTTTCTTTTTCTTTTTGACAGTCTTTTAGATTTAAGAC 110320
Qy      912 ------------------------------------------------------------ 912
Db   110321 TAGAGCTGTGTTACACACAACACCAAAGGCAATTAAAAAAAAAAAAACTCTTTAACTAAA 110380
Qy      912 ------------------------------------------------------------ 912
Db   110381 TTGTAATAAAGAATCTGCTTGTATATGGCTCTATTTGCACCTGGTAGGATAATGCCTAAA 110440
Qy      912 ------------------------------------------------------------ 912
Db   110441 TGTGTCTCTTGTGACTACATGTCTACAGCAACAACTCTTCATCTTATTAGCTTTTCTTCT 110500
Qy      913 ------------------AspGluCysTyrGlnValArgGlnValPheAlaGlnLysLe 926
                              ||||||||||||||||||||||||:::::::||||||||||||
Db   110501 CTTTTCCCTCTCTTCATAGGATGAGTGCTATCAGGTGCGACAGATCTATGCTCAGAAGTT 110560
Qy      926 uHisLysGlyLeuSerArgLeuArgLeuProLeuGluTyrMetAlaIleCysAlaLeuCy 946
            ||||     |||   :::|||   |||||||||||||||||||:::    ||||||||
Db   110561 ACATGTTGCTCTGGTGAAGTTGTTGTTACCACTGGAGTACATGGCTGTATTTGCTTTATG 110620
Qy      946 sAlaLysAspProValLysGluArgArgAlaHisAlaArgGlnCysLeuValLysAsnIl 966
            ||||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db   110621 TGCTAAAGACCCAGTAAAAGAGCGACGTGCTCATGCTCGACAGTGCCTGCTGAAGAACAT 110680
Qy      966 eAsnValArgArgGluTyrLeuLysGlnHis-Ala------------------------ 977
            |:::||||||||||||||||:::|||||:::
Db   110681 CAGTGTCCGCAGAGAGTACATCAAACAGAACCCTATGGCTCATGGTAGATACTTTTGTTT 110740
Qy      977 ------------------------------------------------------------ 977
Db   110741 TTAATTATTTAAAAACATATATGTTTTGTACTGGTATACCACATAACACACCGAGCCAAT 110800
Qy      978 ---------------------------------------------AlaValSerG 981
                                                          :::||||
Db   110801 TTCCTGCGGAGCTGCTTGAATGGTAAAGATCTATCTAACTAATCTTTTTTTGCATTTCAG 110860
Qy      981 luLysLeuLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaH 1001
            |||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db   110861 AAAAGCTGCTGTCTCTCCTTCCAGAATATGTGGTGCCATACATGATCCATCTTCTAGCTC 110920
Qy     1001 isAspProAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGlu---- 1019
            ||||||||||    |||   ||||||:::||||||||||:::|||||||||
Db   110921 ATGATCCAGACCTTACTAAACCACAGGATTTGGAGCAGCTTAGAGATGTCAAAGA-GTAC 110979
Qy     1019 ------------------------------------------------------------ 1019
Db   110980 GAGAATCATATTTTTGCCATTTCTTTTCACTCAAATGGCCTGACTAATACTGCCAATCT 111039
```

```
Qy       1019 ------------------------------------------------------------ 1019

Db     111040 TTATACATTTTTTAAATATAATTTCTACTTGAACAGATAGATCTAGTTTCCCTAATTTAT 111099

Qy       1020 -----------------------------------------------------------C 1020
                                                                         |
Db     111100 GTTATGTGCGTTTTAAAATGACAACTAATTGAAATGAAAAATAATCTAATTTTTACAGGT 111159

Qy       1020 ysLeuTrpPheValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheI 1040
              ||||||||||:::||||||:::||||||    ||||||||||||||||||||:::|||:
Db     111160 GCTTATGGTTCATGCTGGAGGTTTTGATGACAAAAAATGAGAATAACAGTCACTCGTTCC 111219

Qy       1040 leArgLysMetValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaL 1060
              ::|||||||||||||||||||||||||||||||||||||||||    ||||||||    |
Db     111220 TAAGGAAGATGGTAGAAAACATCAAACAAACAAAAGATGCTCAGTGCCCAGATGACCCAA 111279

Qy       1060 ysMetAsnGlu------------------------------------------------- 1063
              ||    ||||||
Db     111280 AGGCAAACGAGGTTTGTCACATTAACACGTTATGCAGATTGTTTATGTATTTCTGAATCC 111339

Qy       1063 ------------------------------------------------------------ 1063

Db     111340 GCTTGTTAAGTGTGACGTCACCCGAAGCGGCTTCCAGGTCCAAGCGCTCAATTCAACTGA 111399

Qy       1063 ------------------------------------------------------------ 1063

Db     111400 ATGGGGAGACTCATGAAATCGTAATAATAAACGTTTTGCAAAGTGATTTAATACTTTCGA 111459

Qy       1063 ------------------------------------------------------------ 1063

Db     111460 AAATCTCGATCGCAGTATATATGTCCATGCCTAATATCCGATGGCCAAAAGTGATTAATT 111519

Qy       1063 ------------------------------------------------------------ 1063

Db     111520 TTTTTATAAATTGTTAATTTTGGTATTTGTTATGCAGCAAGCCCAGAGATTGTTGTGTAC 111579

Qy       1063 ------------------------------------------------------------ 1063

Db     111580 ACTATGATTTTATATAAAATTAACTTTAATGTGTGATAGGAATAAAACGTGATCATAAAC 111639

Qy       1063 ------------------------------------------------------------ 1063

Db     111640 GAATGATTTCTTCACTCAAATGAATGGCGGCTTGGACCCAGAAACAGTATTACATACGTC 111699

Qy       1063 ------------------------------------------------------------ 1063

Db     111700 ACAAACACGTCACCACTTAACAAGCGGATAGGGTTGTTAAACTTTGATATAAATCTTTAC 111759

Qy       1063 ------------------------------------------------------------ 1063

Db     111760 TGAACGAAAATGTGACTACATTTCCCACTTGCTTTTTGAGTACTGCTGAGTGTATTTATG 111819

Qy       1063 ------------------------------------------------------------ 1063

Db     111820 AACACTGCCGATTAGCCATTGTGCTCATGTGGTCTAAATCCTAGTGTTACAGTATTTGTA 111879

Qy       1064 ----------LysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerLysS 1080
                        |||||||||  |||||||||||||||:::   :::|||   :::||||
Db     111880 TTTATTTACAGAAACTATACATTGTATGTGATGTGGCATTATTTGTGATCGCGAACAAGA 111939

Qy       1080 erThrThrTyrSerLeuGluSerProLysAspProValLeuProAlaArgPhePheThrG 1100
              |||||:::        |||:::|||||||||||||||||||||||::::::|||:::|||
Db     111940 GCACTTCATGTCATCTGGACTCTCCAAAAGATCCTGTGCTGCCCAGCAAGTTCTACACCC 111999

Qy       1100 lnProAspLys------------------------------------------------- 1103
              |||||||||
Db     112000 CACCTGATAAGGTAACAACCACTCAGACAACCACAGCACCTCTTTACATGTGACCATTCT 112059

Qy       1104 -----------------------------AsnPheSerAsnThrLysAsnTyrLeuPro 1113
                                           |||   |||   |||   ||||||
Db     112060 ATGAAGCTCGTTCTCTCTTTCTTTCAACAGGAGTTTGTCAACGATAAGGAATATCTGTCA 112119

Qy       1114 ProGlu-MetLysSerPhePheThrProGlyLys-------------------------- 1124
              ||| :::|||   |||           ||||||
Db     112120 GCAGAAGCTAAAATAGTTCTGCAGACAGGGAAAATCCAGGTAGGTTTTATACCCTATATA 112179

Qy       1124 ------------------------------------------------------------ 1124

Db     112180 TGCACATGGCACCACGTAACACATTTGTGTAATTAATCCGTATATAGCCATGGCTAAAGC 112239

Qy       1124 ------------------------------------------------------------ 1124

Db     112240 ATGATTTTTGTTTTTAGTAAATTTTCGTGACATTCATTTTAAAAAGTAAACACATTTTAG 112299

Qy       1124 ------------------------------------------------------------ 1124

Db     112300 GTTGTAAGGACAGGGTAGGAAACCTCTAGGATTTGGGAGTTAATGAATAGGGATATAGCA 112359

Qy       1124 ------------------------------------------------------------ 1124

Db     112360 AAAATTATTTCTTTACAATTGTTTTGAAACCACACAACTAACCAGAAACGGACAGTAAAT 112419

Qy       1125 ------------------------------------------------ProLysThrTh 1128
                                                              ||||||   ||
Db     112420 CCTTGACTGCATTAAATAACGCACCCTTTGTTGCTCATTTCAGCAACCACCAAAGCAGAC 112479

Qy       1128 rAsnValLeuGlyAlaValAsnLysProLeuSerSerAlaGlyLysGlnSerGlnThrLy 1148
              |  ||||||||||||||||||||||||||||:::      ::::::          ||
Db     112480 AGGCGTTCTGGGGGCTGTAAACAAACCTTAACAGTAACAGCTCGCCGACCGTACATTAA 112539

Qy       1148 sSerSerArgMetGluThrValSerAsnAlaSerSerSerSerAsnProSerSerProGl 1168
              |:::          ||||||   |||||||||||:::::::|||   |||||||||||
Db     112540 AACTATCACCTCTGAGACAGGGAGCAATGCAAGCACCAACTCCCAACCCAGCTCACCTGC 112599

Qy       1168 yArgIleLysGlyArgLeu----------------------------------------- 1174
                    |||   |||
Db     112600 AACAAATAAGAGCAGGTTTGTGCTGATCAGCATTTTCTATTTAAGATTCAATGTGATTTA 112659

Qy       1175 --------------------------------------------------AspSerSe 1177
                                                                |||||
Db     112660 TAATGCAGGATCTTCATAAAGGATTTTTTATTTTCTTTCTATTGAAGGGATGTGAGTTC 112719

Qy       1177 rGluMetAspHisSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLy 1197
              ||||:::      |||||||||
Db     112720 TGAGGTGGGGGCAAGGGAGAACGAA----------------------------------- 112744

Qy       1197 sSerAspLysArgAspAspSerAspLeuValArgSerGluLeuGluLysProArgGlyAr 1217
                                                      |||   |||
Db     112745 ----------------------------------------GAGAATCCTGTGATCAC 112761

Qy       1217 gLysLysThrProValThrGluGln---GluGluLysLeuGlyMetAspAspLeuThrLy 1236
               |||         |||   :::         ||||||:::    |||       |||
Db     112762 AAAAGCAGTCAGTGTAAAGAAGGTATGAAATGAAAAGGTCTGTATGTTCACACTGCTATC 112821

Qy       1236 s--------------------------LeuValGlnGluGln------LysProLysGl 1245
                                         |||:::   |||:::      :::|||   ||
Db     112822 ATCCTCAACTCACTCTTATTTGTTTTTGCTCTTGTAGGAGGAAGCTGCTCAGCCCAGCGG 112881

Qy       1245 ySerGlnArgSerArgLysArgGlyHisThrAlaSerGluSerAspGluGlnGlnTrpPr 1265
              |          |||||||||        |||||||:::    |||
Db     112882 G-----------AGGAAGCGTGCTGCTCCTGCCAGTGATGGGACAGAA----------- 112918

Qy       1265 oGluGluLysArgLeuLysGluAspIleLeuGluAsnGluAspGluGlnAsnSerProPr 1285
                                    :::      |||            :::    |||||
Db     112919 ------------------AACAGTGTGTCTTCAAACCCCTCTGCAGGTTCACAGCCACC 112959

Qy       1285 o---LysLysGlyLysArgGlyArgProProLysProLeuGlyGlyGlyThrProLysGl 1304
              |      |||    :::|||||||||||||||||||       ||||||       :::||
Db     112960 GCTCAATAAACCACGACGCGGTCGACCCCCTAAAAACAGTGCGGGCGGTGCCACACAGGA 113019

Qy       1304 uGluProThrMetLysThrSerLysLysGlySer---------LysLysLysSerGlyPr 1321
              |:::            |||:::      |||:::         :::|||:::::   ||
Db     113020 AAAAGAGGCTGGAGCTACGACAGGGGCAGGAGCCGGACGGGGCCGGAAGAGAGCAGCACC 113079

Qy       1321 oProAlaProGluGluGlu-GluGluGluGluArgGlnSerGlyAsnThrGluGlnLysS 1341
              |                    :::   :::         |||          |||   |
```

```
Db      113080 TTCTCAAGATCCTGCCAGCACAGCCTCAACTGATGCACTCAGTGACAAGACACCAAAACA 113139

Qy        1341 erLysSerLysGlnHisArgValSerArgArgAlaGlnGlnArg 1355
               ||:::  ||||||  |||      :::|||   :::   |||
Db      113140 GCAGAAAGAAGCAGAGCCGAAAAGAGCAGCGCCACAGAGACAGA 113183

RESULT 14
AC007475
LOCUS       AC007475              173613 bp    DNA     linear   INV 28-FEB-2001
DEFINITION  Drosophila melanogaster, chromosome 2R, region 49A-49B, BAC clone
            BACR04E21, complete sequence.
ACCESSION   AC007475
VERSION     AC007475.7  GI:13162478
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 173613)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 2R, region 49A-49B
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173613)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Feb 28, 2001 this sequence version replaced gi:6957978.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .173613
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
                     /map="49A-49B"
                     /clone="BACR04E21 (D592)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT     48071 a  37844 c  38273 g  49425 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                2.24e-80        Length:        173613
Score:                    1729.50         Matches:       463
Percent Similarity:       49.43%          Conservative:  270
Best Local Similarity:    31.22%          Mismatches:    436
Query Match:              24.04%          Indels:        318
DB:                       3               Gaps:          37

US-09-512-581B-2 (1-1391) x AC007475 (1-173613)

Qy          13 IleThrTyrProProGlyValLysGluIleSerAspLysIleSerLysGluGluMetVal 32
               |||  ||||||  |||   :::   :::   :::   :::         :::|||:::::
Db       22028 ATAGTGTACCCCACCGGATGCCGGCCGCTGGTCGAGGATTTGGGCACAGACGAGCTGATA 22087

Qy          33 ArgArgLeuLysMet-ValVal------------------------------------- 39
               ||||||||||||::: |||:::
Db       22088 AGGCGCCTCAAGGTGAGTCCTGCCCCTGGTTCCCTTAGCAAAACTTCTCACTTAGAATGA 22147

Qy          40 ---------------------LysThrPheMetAspMetAspGlnAspSerGluGluGl 52
                                              |||||||||||
Db       22148 TTCTGTTTTACAGACCCTCGCCAATGTTCTGCAGACCATGGACCAGGACGACAATCTCTA 22207

Qy          52 uLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPhePheLeuLysHisPr 72
                ::::::   |||:::  ||||||||||||||       |||||||||:::::||||
Db       22208 CCAGCAA---TACATACCATTGGCCCTCCATCTGCTGGACGACTTCTTTATGCAACATCC 22264

Qy          72 oGlyLysAspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAl 92
               |   :::||||||:::|||||||:::|||||||||:::||||||:::    |||:::||||
Db       22265 ATCCCGTGATGTCCAGCTCTTAATCGCGTGCTGCGTGGCGGATGTCCTGCGGGTTTATGC 22324

Qy          92 aProGluAlaProTyrThrSerProAspLysLeuLysAspIlePheMetPheIleThrAr 112
               |||||||||||||||           |||:::::|||   ||||||   |||       ::
Db       22325 GCCAGAGGCTCCCTACAAAGAGCAAGATCAAATCAAAACTATATTTAAGTTTTTATCAA 22384

Qy         112 gGlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLe 132
               :||||||   ||||||:::|||   :::    |||    |||    |||||||||||||||
Db       22385 GCAACTGCACGGTCTGAAGGATCCGCGCGATCCCTCTTTCAAGCGTTACTTCTATCTGCT 22444

Qy         132 uGluAsnIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSerAsnGl 152
               ||||||:::|||:::|||||||||:::|||:::|||||||||||||||||       ||
Db       22445 GGAGAACCTTGCTTTTGTCAAGTCTTTTAATATGTGTTTCGAACTGGAGGACTGCCAAGA 22504

Qy         152 uIlePheThrGlnLeuTyrArgThrLeuPheSerValIle------------------- 165
               |||||||      |||:::   |||:::|||    ::::::
Db       22505 AATCTTCCAAGACCTGTTTAGTACCATCTTCAAGATTGTCAAGTAAGTCGAGCTCTTCCT 22564

Qy         166 -----------------------------------------AsnAsnGlyHisAsnGl 171
                                                        ::::::   |||:::
Db       22565 TTATCCTGCGAACTCTTTATAATCACATGTTATATAATTTTCTAGTGACCAGCACAGTGT 22624

Qy         171 nLysValHisMetHisMetValAspLeuMetSerSerIleIleCysGluGlyAspThrVa 191
               ||||||              :::|||::::::|||    :::|||   |||   |||    ::
Db       22625 CAAGGTGACTAACTTTTTTCTAGATGTGCTAAGCCCGTTGATCACAGAAGCCGACAATCT 22684

Qy         191 lSerGlnGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAs 211
               :|||   ||||||||||||   :::|||:::|||:::|||          :::|||:::   ||
Db       22685 GTCGGTGGAATTGTTGGATCTCATTCTGATTAACATCGTAGAGCCGTACAAATCTAACAA 22744

Qy         211 nLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGlu-P 231
               ||||   |||         |||   :::   ||||||   :::|||        |||:::||| |
Db       22745 CAAGTTTGCCTGCCAATTAACAGAACAACTTCTCACTAAAACGGGCGATGCTCTCGAGTC 22804

Qy         231 ro---------------------------TyrIleThrThr----------------- 235
               ||                           :::   |||||
Db       22805 CACCATCAAAATGGTAATATTTCGTTGATCTTTGGTGCACCACGTTATTCAACCTAATCC 22864

Qy         236 -----------PhePheAsnGlnValLeuMetLeuGlyLys------ThrSerIleSerA 250
                          ||||||||:::   |||::::::   |||         |||   ::::|||
Db       22865 CTGATCTACAGTTCTTTAACCGTGCTCTGGTCATGGATAAGCCAAACACAAAACTGTCC- 22923

Qy         250 spLeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuL 270
                 ::::::          ::::::|||:::|||   ||||||        |||::::::   ||||
```

```
Db      22924 --ATTACAAACAAAATCTATGATATCATTTACGAGCTTAACCGCATCAATGCAGGCTTGT 22981

Qy        270 euLeuSerValLeuProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluGluArgL 290
              ||   ||||||||||||||||||||   ||||||   |||   |||:::   |||||||
Db      22982 TGTGTTCAGTGCTACCTCAGCTGGAGAACAAACTGCTGTCCACGGACGATGCTGAGAGGC 23041

Qy        290 eu---------------------------------------------------------- 290
              ||
Db      23042 TAAGTGAGTAATTTATCAGCATGTACCTCTATTTGCATATTAATAAACATATTTTCGTTA 23101

Qy        291 -----GlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaS 309
                   :::            ||||||::::::||||||         |||||||||:::||||||
Db      23102 CTACAGAGGCAACCACTCTCTTGTCTCGTATGTTCTCCGAAAAGGACTCGCAACTAGCGA 23161

Qy        309 erGlnAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProI 329
                :::         |||   :::   :::   ||||||||   ||||||       |||:
Db      23162 AAAAGTACCCCAACCTGTTGAAAATCTTCTTTGGGCGCTTCTGCGACATCACCGAACCAG 23221

Qy        329 leArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaL 349
              ::|||::::::||||||:::      :::   |||   |||:::|||||||||    |||
Db      23222 TTCGCATCAAGTGTGTTCAGTCATCCATGCACTTCCTGCTCAATCACCCAAGTCTTCAGC 23281

Qy        349 ysAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleArgHisA 369
                |||::::||||||   |||::::::|||::::||||||      :::|||   :::||||||:
Db      23282 ACGATATCACCGAGAAATTGCGGCTTCGAAATCACGATCTTGACGAAGTGGTCCGCCACG 23341

Qy        369 spValIleValSerIleValThrAlaAlaLysLysAspIleLeuLeuValAsnAsp---- 387
              ::|||:::::::::::::||||||           ||||||:::|||   ||||||   :::
Db      23342 AGGTGGTAATGGCTATTGTGGAAACTGCCAAGCGCGACTTCACCCTCGTTCTCGAAGCGC 23401

Qy        388 --HisLeuLeuAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLysG 407
                    ||||||    |||||||||||||||||||||||||::::::::::::|||::::
Db      23402 CCGATCTACTAGAAATTGTGCGCGAGCGGACGCTAGATAAGAAGTACAAAATTCGCAGGG 23461

Qy        407 luAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyL 427
              ::||||||   |||||||||   ||||||||||:::          :::
Db      23462 ATGCGATGAATGGTCTGGCCTACATCTACAAGCGCGCAATTTGCGAACCCAACGATCTAA 23521

Qy        427 ysAspAlaAlaLys---GlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrG 446
                          |||   ::::::   ||||||||:::|||:::||||||   ||||||:
Db      23522 GCACTGGCCTTAAGGTCAGGGTTGACTGGATTAAGAACAAGATACTACATGGATACTACA 23581

Qy        446 lnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProH 466
              ::     :::::|||||||||||||||||||||||:::          :::||||||:
Db      23582 AAGTGGGCTTGGAGGACCGCCTGCTTGTGGAACGCCTACTTATCACCTGCCTGGTTCCCT 23641

Qy        466 isAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspL 486
              ::   |||         |||||||||||   ||||||:::|||         ||||||
Db      23642 ATAAACTAGCTCCGGAAGAGCGCATGAAGAAGCTATATCATTTGCTAGGCGATCTCGATG 23701

Qy        486 euAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnV 506
                ||||||   ||||||      |||:::   |||   |||         |||:::   |
Db      23702 CCAATGCCACCAAGGCTTTCGTCGAGCTGCAGAAGAACCAAATGAAGACACGCAATACGG 23761

Qy        506 alLysAspLeuLeuAspLeuIleLysGln-----------ProLysThrAspAlaSerV 522
              ||   |||   :::   |||                     |||:::      :::   :
Db      23762 TCAGTGATTGGATCAAGCTACACCACTCTAAGGAGTTCACGCCGCGCGTACTTAGTCAGC 23821

Qy        522 alLysAla---------------------------------------------------- 524
              ::   |||
Db      23822 TCAGCGCCAAGCAGGCCAACATTGCCAAGTGGGTTCTATAATAAGTTTTCTTTTGATAGT 23881

Qy        525 --IlePheSerLysValMetValIleThr-ArgAsnLeuProAspProGlyLysAlaGln 543
                :::   |||   |||                |||   ||||||||||||   ||||||
Db      23882 TTGTAACCAGTGATGTGTTTTTTTTTCTTCAGACTGCTTCCAGATCCCCTAAAAGCGGCA 23941

Qy        544 AspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeu 563
              :::::::::   :::|||:::         |||      |||   ::::::   :::   :::
Db      23942 GAGTACCTAACCCAGTTTAGCAACAACCTGCGAAAAGATGCTCAGCTCCTACGATGCATT 24001

Qy        564 GluValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIle 583
                 :::::::::               ||||||::::::               :::      :::
Db      24002 AACATTGTCCTTAAACGCGACGTAAGCTGTCGGGAGTGTGCTGACACGATGGGAGTTCTC 24061

Qy        584 ThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPhe 603
                ||||||||||||            |||      :::|||      :::         :::|||
Db      24062 CTGAAAAAGCTTGGCGCCCACGTCCAA---TCGAATCTGTATTACAACACAGTTAAGATG 24118

Qy        604 LeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLys 623
              |||::::||||||:::|||   |||      :::|||   |||||||||      ||||||:::
Db      24119 CTGATTGAGCGCGTGGCATCAGTGATGGTGGACAAGGAGTCTATTGGCGTACTTATTAGG 24178

Qy        624 ------------------------------------------------------Gln 624

Db      24179 TAAATTGCTTTTAAGTGCCCTTTTGCCAGACTTAATCCGGTAATCTTTTTGCTCAGCCTA 24238

Qy        625 ValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGln 644
              :::   :::   |||:::         :::         |||:::   |||:::         :::
Db      24239 ATTGAACAATGCATCGAGAAAGGGTCCATGTGTGAAGAGATCGGAATTTCGGCTCAGGAA 24298

Qy        645 AlaIleArgAlaGlyLeuGluLeuLeu------------------------------- 653
              |||         ||||||:::||||||
Db      24299 GCGGGCGAGCGCGGACTCAAGTTGTTGGCCGTGAGTATAGAGATAATTAAGTTATGATA 24358

Qy        654 ----------------------------LysValLeuSerPheThrHisProIleSer 663
                                          :::   ||||||:::
Db      24359 CAGAAATCCATTTGATTGCTTTATACTTGCAGATG-CTTTCATACGTCTTTTCGGCTCAT 24417

Qy        664 PheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLys 683
              |||   :::      :::         |||:::::::   |||         :::   :::
Db      24418 TTCTTCACCGATACTTCGCTGCGTCATTTAATTTCCCTGCTCAGTTACGAGCAGGATTAT 24477

Qy        684 ValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGly---------SerLysIleGlu 700
              ||||||         |||:::                                |||:::
Db      24478 GTTGCACCACTGGTGCTAAAGACACTAACTCATCTGGGGCGCTATCAGCCACTCATTGAC 24537

Qy        701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLys 720
              :::      |||   |||         |||   |||||                 :::
Db      24538 GACCCTACACCGGCTATTCTCGACGAGCTGGCTCCGGTGTGCAAGGACTTCGCACTTATT 24597

Qy        721 GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHis------------------ 734
              |||   |||:::|||||||||:::|||:::      ||||||
Db      24598 GGAACCCCCAAGCAAGCAAAGCACGCGGTGCGATGTATTTCGTAAACAGCCAGTCGTCG 24657

Qy        735 --------------------------AlaIlePheSerSerLysGluThrGln------ 743
                                        |||      :::|||      |||||
Db      24658 GCTTCCACTGATGGAGCAACAAGTGGAGCTGGAAGTGCGTCCACAACTACACAAACAGTG 24717

Qy        744 ---------PheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
                       |||   :::|||   |||   |||         |||   |||   |||
Db      24718 CATCCTATTTTCAACGAGATCATCGAGACGCTTCGCCTAAAACTGACACCA---AACTGT 24774

Qy        761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
              ||||||      |||   :::||||||:::||||||||||||            |||
Db      24775 GAGCATCAGCGCACAAAGATTGTGACCTTGGGTCACATTGCCTACAATATGCCACAGGCC 24834

Qy        781 PheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
              |||      |||   |||:::      :::|||         |||||||||:::||||||:::
Db      24835 TTCCTAACGCCCATTAAGAATATGATTGCGCGACGAATTGTCAAAGAGCTGCTTATCCAA 24894

Qy        801 AspArgLeuProGlyLysLysThrThrLysLeu--------------TrpValProAsp 815
              :::   :::|||   ::::::         :::|||                  |||
Db      24895 GAA---GTTCCTGCGCAGCGAGACTACGAACTGCCAGAGGACAGTGATTGGTGTGCCCAA 24951

Qy        816 GluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrp 835
              |||::::::   |||:::|||:::   |||:::   |||:::|||   |||   ||||||
Db      24952 GAGAAACTACCGCCGGACACTCTATGTAAGCTGGATGCGCTCAAGGCTATGGCCAGGTGG 25011

Qy        836 LeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThr 855
              ||||||||||::::::   :::                  :::         |||   |||:::|||
Db      25012 CTTTTGGGATTGCGTACCGAT---GAGCACGCTGCCCAGAAAACATTCCGAATGCTGGCG 25068

Qy        856 ThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMet 875
                 ::::::         |||||||||      |||      ::::::               :::
Db      25069 GCGTTTGTTAACCAACGAGGCGATTTGCTCGGCCAGAACCGTCTTTGCGGTGCCGAGAAA 25128
```

```
Qy        876 SerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyr 895
              |||  |||||||||  |||      |||::::::|||:::   ::::::
Db      25129 TCTTGGCTGCGCCTCGGGGCAGCCTGCGCCATGCTCAAAGTGTGCGAACAAAAGGGCGTA 25188

Qy        896 HisGluIleIleThrLeuGluGlnTyr---GlnLeuCysAlaLeuAlaIleAsnAspGlu 914
               :::      :::   |||||||||  ||||||        |||   ::::::   :::
Db      25189 GGTGATCAGTACAGCGCTGAGCAGTATTTGCAGCTTTCCCAGCTGATGGTGAGTAGAGAC 25248

Qy        914 ------------------------------------------------------------ 914

Db      25249 ATTCATCTTGTGCTTAGAACTTAGTCTCAATGTCTATCTCGGTGTTCCCGTAGGCTGATC 25308

Qy        915 ---CysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeu 933
                 |||   :::||||||::::::||||||:::|||||||||||||||||||||
Db      25309 CGGTGCCA-GAAGTTCGGGAAATCTTTGCTCGCAAGCTGCACAAAGGATTAAGCAGAAGT 25367

Qy        934 Arg------------LeuProLeuGluTyrMetAlaIleCysAlaLeuCysAla------ 947
                             |||||||||::::::|||   :::     |||
Db      25368 TTGCCCAGGAACTGTTTGCCGCTGGACTTCATGGGCTTGTATGTGCTGGCTGGTCTAGAG 25427

Qy        948 ---------LysAspProValLysGluArgArgAlaHisAlaArg-GlnCysLeuValLy 964
                       |||      |||:::|||                 |||  :::    |||:::
Db      25428 ACTGAGAGGAAGTGAGTAGTTCAAGAAACCATCCGATCCGTTCGGAGAAGTATTGATAAA 25487

Qy        964 s-------------------------------------------AsnIleAsnValAr 969
                                                          ::::::|||   ||
Db      25488 TATTTTCAACAGATTGCAAGACCTTGTGCGTCACTATGCAGAAACGGATGTAAACAAACG 25547

Qy        969 gArgGluTyrLeuLys------------------------------------------- 974
              |||||||||||||||
Db      25548 GCGGGAATATCTCAAGACTGTGGCTATGACATGTAGGTGTTTTACTTTTGGAATCATATT 25607

Qy        975 ---------------------------------GlnHis-AlaAlaValSerGluLysL 983
                                                 ::::::           :::
Db      25608 GTGTATTTTAACCATTAAATCAAATTATTCCAGCTCCCGACAGCTCAACGGAGTCACAAT 25667

Qy        983 euLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspP 1003
                |||   :::||||||:::|||::::::   :::    |||     :::|||   ||||||
Db      25668 CATTACACATACTACCTGACTACATGCTGGCTTTCGCTATTCCCGTGCTGGTCCACGATC 25727

Qy       1003 roAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpP 1023
               ||   :::           :::|||     ||||||:::   :::::::::||||||   |
Db      25728 CACGCTTCACGAATCACGAGGACTACGTACAGCTGCGCAAGATGGAGAAGTGCCTGCGTT 25787

Qy       1023 heValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysM 1043
              |||:::||||||   ||||||||||||  |||         |||:::|||  :::::::
Db      25788 TCATTCTGGAGCCGCTGATGGCCAAACGAGAAACGTTTGTCCATAGCTTCTACAAGCAGC 25847

Qy       1043 etValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsn- 1062
              ::::::::    ||||||       :::   :::   |||      |||   :::   |||
Db      25848 TGCTGCAGCTGATAAAGCATCGCGAGTTCAGTCTGGGG---TCGGACAAGCGCGACAACT 25904

Qy       1062 ------------------------------------------------------------ 1062

Db      25905 ATGTATGTCACTCAATCATTTTCCCAGGCTTTCCCTTCTAAATTATATTTATACATATAT 25964

Qy       1063 ----------GluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerL 1079
                        :::|||::::::   :::||||||:::||||||    ||||||    ||||
Db      25965 TTTTCGGTTATCAGAAAATGTGGGCGCTCTGCGATCTTGCCATGTACATTATCGACTCCA 26024

Qy       1079 ys--------------------SerThrThrTyrSerLeuGluSerProLysAspProV 1092
               ||                   ::::::|||:::|||:::
Db      26025 AGTTCAGTCCATTTGATGGCAACACGAGCACCTTTTCCATGCCGCTG-----------G 26072

Qy       1092 alLeuProAlaArgPhePheThrGlnProAsp---LysAsnPheSerAsnThrLysAsnT 1111
                ||||||      ::::::   :::|||          ||||||   |||           |
Db      26073 CTTTGCCAGAAATGTATTATAAAGAGCCTGCCGTTGCGAATTTCCAAAACAATGACGTCT 26132

Qy       1111 yrLeuProProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValL 1131
              |||:::|||   ::::::           :::|||   |||           |||:::
Db      26133 ATATACCGCTGGACGTG---------TATACGCTGGGAGCCAAATCCACGAGCAAAGCTG 26183

Qy       1131 euGlyAlaValAsnLysProLeu---SerSerAlaGlyLysGlnSerGlnThrLysSerS 1150
```

```
                |||        :::    ||||||:::|||    |||          ::::::|
Db      26184 CCGCAACAGC-AATGACAACGTCGCGAGCAGCAGTGGCTC-CAAAGAGACCGGCGGAACA 26241

Qy       1150 erArg--------------------------MetGluThrValSerAsnAlaSerSerS 1161
               ||:::.                         :::   ::::::   |||      :::
Db      26242 GTCAATCATGGATGATGAAAATCCGCAGGTGAGTGAGGTCCTTGGAAAATACTGAGAATG 26301

Qy       1161 erSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspH 1181
                ||||||                    :::       :::    |||:::
Db      26302 AATCTAACTAGTGTGTGAAATTTGATTTCAGGAGAACAATCTGTTCGACAACATAC---- 26357

Qy       1181 isSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
                        |||     |||:::   ||||||   |||    ::::::::   :::
Db      26358 ----------GAGCGGCTGACACTACGGAGCCCATGGCCAAACGAACGCGCGCAGGAG 26405
```

RESULT 15
AC007474

```
LOCUS       AC007474             192763 bp    DNA     linear   INV 28-FEB-2001
DEFINITION  Drosophila melanogaster, chromosome 2R, region 49A-49B, BAC clone
            BACR14L19, complete sequence.
ACCESSION   AC007474
VERSION     AC007474.5  GI:13162477
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 192763)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D.; Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 2R, region 49A-49B
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 192763)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Feb 28, 2001 this sequence version replaced gi:5670617.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1..192763
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /strain="y; cn bw sp"
```

```
                  /db_xref="taxon:7227"
                  /chromosome="2R"
                  /map="49A-49B"
                  /clone="BACR14L19 (D591)"
                  /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                  Drosophila melanogaster BAC library, partial EcoRI in
                  pBACe3.6)"
BASE COUNT     54505 a  41459 c  42141 g  54658 t
ORIGIN

Alignment Scores:
Pred. No.:              2.56e-80        Length:         192763
Score:                  1729.50         Matches:        463
Percent Similarity:     49.43%          Conservative:   270
Best Local Similarity:  31.22%          Mismatches:     436
Query Match:            24.04%          Indels:         318
DB:                     3               Gaps:           37

US-09-512-581B-2 (1-1391) x AC007474 (1-192763)

Qy         13 IleThrTyrProProGlyValLysGluIleSerAspLysIleSerLysGluGluMetVal 32
              |||   ||||||   |||   :::   :::   :::   :::    :::|||::::::
Db     146305 ATAGTGTACCCCACCGGATGCCGGCCGCTGGTCGAGGATTTGGGCACAGACGAGCTGATA 146364

Qy         33 ArgArgLeuLysMet-ValVal------------------------------------- 39
              ||||||||||||::: |||:::
Db     146365 AGGCGCCTCAAGGTGAGTCCTGCCCCTGGTTCCCTTAGCAAAACTTCTCACTTAGAATGA 146424

Qy         40 ---------------------LysThrPheMetAspMetAspGlnAspSerGluGluGl 52
                                               ||||||||||||
Db     146425 TTCTGTTTTACAGACCCTCGCCAATGTTCTGCAGACCATGGACCAGGACGACAATCTCTA 146484

Qy         52 uLysGluLeuTyrLeuAsnLeuAlaLeuHisLeuAlaSerAspPhePheLeuLysHisPr 72
               ::::::   |||:::   ||||||||||||||    ||||||||||:::::||||||
Db     146485 CCAGCAA---TACATACCATTGGCCCTCCATCTGCTGGACGACTTCTTTATGCAACATCC 146541

Qy         72 oGlyLysAspValArgLeuLeuValAlaCysCysLeuAlaAspIlePheArgIleTyrAl 92
              |   :::||||||:::|||||||:::|||||||||:::|||||||:::    |||:::|||||
Db     146542 ATCCCGTGATGTCCAGCTCTTAATCGCGTGCTGCGTGGCGGATGTCCTGCGGGTTTATGC 146601

Qy         92 aProGluAlaProTyrThrSerProAspLysLeuLysAspIlePheMetPheIleThrAr 112
              |||||||||||||||          |||::::::|||   ||||||   |||       ::
Db     146602 GCCAGAGGCTCCCTACAAAGAGCAAGATCAAATCAAAACTATATTTAAGTTTTTATCAA 146661

Qy        112 gGlnLeuLysGlyLeuGluAspThrLysSerProGlnPheAsnArgTyrPheTyrLeuLe 132
              :||||||   |||||||:::|||   :::   |||   |||   ||||||||||||||||||
Db     146662 GCAACTGCACGGTCTGAAGGATCCGCGCGATCCCTCTTTCAAGCGTTACTTCTATCTGCT 146721

Qy        132 uGluAsnIleAlaTrpValLysSerTyrAsnIleCysPheGluLeuGluAspSerAsnGl 152
              |||||||:::|||:::|||||||:::|||||||||||||||||||||         ||
Db     146722 GGAGAACCTTGCTTTTGTCAAGTCTTTTAATATGTGTTTCGAACTGGAGGACTGCCAAGA 146781

Qy        152 uIlePheThrGlnLeuTyrArgThrLeuPheSerValIle-------------------- 165
              ||||||      |||:::   |||:::|||   ::::::
Db     146782 AATCTTCCAAGACCTGTTTAGTACCATCTTCAAGATTGTCAAGTAAGTCGAGCTCTTCCT 146841

Qy        166 ------------------------------------------AsnAsnGlyHisAsnGl 171
                                                         ::::::   |||:::
Db     146842 TTATCCTGCGAACTCTTTATAATCACATGTTATATAATTTTCTAGTGACCAGCACAGTGT 146901

Qy        171 nLysValHisMetHisMetValAspLeuMetSerSerIleIleCysGluGlyAspThrVa 191
               ||||||               :::|||:::::::|||   :::|||   |||   |||   ::
Db     146902 CAAGGTGACTAACTTTTTTCTAGATGTGCTAAGCCCGTTGATCACAGAAGCCGACAATCT 146961

Qy        191 lSerGlnGluLeuLeuAspThrValLeuValAsnLeuValProAlaHisLysAsnLeuAs 211
              :|||   ||||||||||||   :::|||:::|||:::|||         :::|||:::   ||
Db     146962 GTCGGTGGAATTGTTGGATCTCATTCTGATTAACATCGTAGAGCCGTACAAATCTAACAA 147021

Qy        211 nLysGlnAlaTyrAspLeuAlaLysAlaLeuLeuLysArgThrAlaGlnAlaIleGlu-P 231
              ||||   |||         |||   :::   ||||||   :::|||       |||::|||   |
Db     147022 CAAGTTTGCCTGCCAATTAACAGAACAACTTCTCACTAAAACGGGCGATGCTCTCGAGTC 147081

Qy        231 ro-----------------------------TyrIleThrThr---------------- 235
              ||                                       :::   ||||||
Db     147082 CACCATCAAAATGGTAATATTTCGTTGATCTTTGGTGCACCACGTTATTCAACCTAATCC 147141

Qy        236 -----------PhePheAsnGlnValLeuMetLeuGlyLys------ThrSerIleSerA 250
                         |||||||||:::   |||:::::   |||         |||   :::|||
Db     147142 CTGATCTACAGTTCTTTAACCGTGCTCTGGTCATGGATAAGCCAAACACAAAACTGTCC- 147200

Qy        250 spLeuSerGluHisValPheAspLeuIleLeuGluLeuTyrAsnIleAspSerHisLeuL 270
                ::::::      :::::|||:::|||   ||||||      |||::::::   ||||
Db     147201 --ATTACAAACAAAATCTATGATATCATTTACGAGCTTAACCGCATCAATGCAGGCTTGT 147258

Qy        270 euLeuSerValLeuProGlnLeuGluPheLysLeuLysSerAsnAspAsnGluGluArgL 290
              ||   |||||||||||||||||||   ||||||   |||   |||:::   |||||||
Db     147259 TGTGTTCAGTGCTACCTCAGCTGGAGAACAAACTGCTGTCCACGGACGATGCTGAGAGGC 147318

Qy        290 eu---------------------------------------------------------- 290
              ||
Db     147319 TAAGTGAGTAATTTATCAGCATGTACCTCTATTTGCATATTAATAAACATATTTCGTTA 147378

Qy        291 -----GlnValValLysLeuLeuAlaLysMetPheGlyAlaLysAspSerGluLeuAlaS 309
                   :::         ||||||:::::::||||||      |||||||||:::|||||||
Db     147379 CTACAGAGGCAACCACTCTCTTGTCTCGTATGTTCTCCGAAAAGGACTCGCAACTAGCGA 147438

Qy        309 erGlnAsnLysProLeuTrpGlnCysTyrLeuGlyArgPheAsnAspIleHisValProI 329
                :::         |||   :::   :::   |||||||||   ||||||      |||:
Db     147439 AAAAGTACCCCAACCTGTTGAAAATCTTCTTTGGGCGCTTCTGCGACATCACCGAACCAG 147498

Qy        329 leArgLeuGluCysValLysPheAlaSerHisCysLeuMetAsnHisProAspLeuAlaL 349
              ::|||:::::::||||||:::   :::   |||   |||::::||||||||   |||
Db     147499 TTCGCATCAAGTGTGTTCAGTCATCCATGCACTTCCTGCTCAATCACCCAAGTCTTCAGC 147558

Qy        349 ysAspLeuThrGluTyrLeuLysValArgSerHisAspProGluGluAlaIleArgHisA 369
                |||:::||||||   |||::::::|||::::||||||      :::|||   :::|||||:
Db     147559 ACGATATCACCGAGAAATTGCGGCTTCGAAATCACGATCTTGACGAAGTGGTCCGCCACG 147618

Qy        369 spValIleValSerIleValThrAlaAlaLysLysAspIleLeuLeuValAsnAsp---- 387
              ::|||::::::::::||||||         ||||||:::|||         ||||||    :::
Db     147619 AGGTGGTAATGGCTATTGTGGAAACTGCCAAGCGCGACTTCACCCTCGTTCTCGAAGCGC 147678

Qy        388 --HisLeuLeuAsnPheValArgGluArgThrLeuAspLysArgTrpArgValArgLysG 407
                  ||||||      ||||||||||||||||||||||:::::::::::::::||||:::
Db     147679 CCGATCTACTAGAAATTGTGCGCGAGCGGACGCTAGATAAGAAGTACAAAATTCGCAGGG 147738

Qy        407 luAlaMetMetGlyLeuAlaGlnIleTyrLysLysTyrAlaLeuGlnSerAlaAlaGlyL 427
              ::||||||      |||||||||      |||||||||:::               :::
Db     147739 ATGCGATGAATGGTCTGGCCTACATCTACAAGCGCGCAATTTGCGAACCCAACGATCTAA 147798

Qy        427 ysAspAlaAlaLys---GlnIleAlaTrpIleLysAspLysLeuLeuHisIleTyrTyrG 446
                           |||   ::::::   |||||||||:::|||:::||||||      ||||||:
Db     147799 GCACTGGCCTTAAGGTCAGGGTTGACTGGATTAAGAACAAGATACTACATGGATACTACA 147858

Qy        446 lnAsnSerIleAspAspArgLeuLeuValGluArgIlePheAlaGlnTyrMetValProH 466
              ::        ::::::|||||||||||||||||||||:::                 :::||||||:
Db     147859 AAGTGGGCTTGGAGGACCGCCTGCTTGTGGAACGCCTACTTATCACCTGCCTGGTTCCCT 147918

Qy        466 isAsnLeuGluThrThrGluArgMetLysCysLeuTyrTyrLeuTyrAlaThrLeuAspL 486
              ::   |||         ||||||||||||   ||||||::::|||         ||||||
Db     147919 ATAAACTAGCTCCGGAAGAGCGCATGAAGAAGCTATATCATTTGCTAGGCGATCTCGATG 147978

Qy        486 euAsnAlaValLysAlaLeuAsnGluMetTrpLysCysGlnAsnLeuLeuArgHisGlnV 506
                 ||||||   ||||||         |||:::   |||   |||         |||:::   |
Db     147979 CCAATGCCACCAAGGCTTTCGTCGAGCTGCAGAAGAACCAAATGAAGACACGCAATACGG 148038

Qy        506 alLysAspLeuLeuAspLeuIleLysGln-----------ProLysThrAspAlaSerV 522
              ||   |||   :::   |||                    |||:::         :::   :
Db     148039 TCAGTGATTGGATCAAGCTACACCACTCTAAGGAGTTCACGCCCGCGCGTACTTAGTCAGC 148098

Qy        522 alLysAla---------------------------------------------------- 524
              ::   |||
Db     148099 TCAGCGCCAAGCAGGCCAACATTGCCAAGTGGGTTCTATAATAAGTTTTCTTTTGATAGT 148158

Qy        525 --IlePheSerLysValMetValIleThr-ArgAsnLeuProAspProGlyLysAlaGln 543
                :::   |||   |||                   |||   ||||||||||||   ||||||
```

```
Db   148159 TTGTAACCAGTGATGTGTTTTTTTTCTTCAGACTGCTTCCAGATCCCCTAAAAGCGGCA 148218

Qy      544 AspPheMetLysLysPheThrGlnValLeuGluAspAspGluLysIleArgLysGlnLeu 563
Db   148219 GAGTACCTAACCCAGTTTAGCAACAACCTGCGAAAAGATGCTCAGCTCCTACGATGCATT 148278

Qy      564 GluValLeuValSerProThrCysSerCysLysGlnAlaGluGlyCysValArgGluIle 583
Db   148279 AACATTGTCCTTAAACGCGACGTAAGCTGTCGGGAGTGTGCTGACACGATGGGAGTTCTC 148338

Qy      584 ThrLysLysLeuGlyAsnProLysGlnProThrAsnProPheLeuGluMetIleLysPhe 603
Db   148339 CTGAAAAAGCTTGGCGCCCACGTCCAA---TCGAATCTGTATTACAACACAGTTAAGATG 148395

Qy      604 LeuLeuGluArgIleAlaProValHisIleAspThrGluSerIleSerAlaLeuIleLys 623
Db   148396 CTGATTGAGCGCGTGGCATCAGTGATGGTGGACAAGGAGTCTATTGGCGTACTTATTAGG 148455

Qy      624 -------------------------------------------------------Gln 624
Db   148456 TAAATTGCTTTTAAGTGCCCTTTTGCCAGACTTAATCCGGTAATCTTTTTGCTCAGCCTA 148515

Qy      625 ValAsnLysSerIleAspGlyThrAlaAspAspGluAspGluGlyValProThrAspGln 644
Db   148516 ATTGAACAATGCATCGAGAAAGGGTCCATGTGTGAAGAGATCGGAATTTCGGCTCAGGAA 148575

Qy      645 AlaIleArgAlaGlyLeuGluLeuLeu--------------------------------- 653
Db   148576 GCGGGCGAGCGCGGACTCAAGTTGTTGGCCGTGAGTATAGAGATAATTTAAGTTATGATA 148635

Qy      654 -----------------------------LysValLeuSerPheThrHisProIleSer 663
Db   148636 CAGAAATCCATTTGATTGCTTTATACTTGCAGATG-CTTTCATACGTCTTTTCGGCTCAT 148694

Qy      664 PheHisSerAlaGluThrPheGluSerLeuLeuAlaCysLeuLysMetAspAspGluLys 683
Db   148695 TTCTTCACCGATACTTCGCTGCGTCATTTAATTTCCCTGCTCAGTTACGAGCAGGATTAT 148754

Qy      684 ValAlaGluAlaAlaLeuGlnIlePheLysAsnThrGly---------SerLysIleGlu 700
Db   148755 GTTGCACCACTGGTGCTAAAGACACTAACTCATCTGGGGCGCTATCAGCCACTCATTGAC 148814

Qy      701 GluAspPheProHisIleArgSerAlaLeuLeuProValLeuHisHisLysSerLysLys 720
Db   148815 GACCCTACACCGGCTATTCTCGACGAGCTGGCTCCGGTGTGCAAGGACTTCGCACTTATT 148874

Qy      721 GlyProProArgGlnAlaLysTyrAlaIleHisCysIleHis------------------ 734
Db   148875 GGAACCCCCAAGCAAGCAAAGCACGCGGTGCGATGTATTTCGTAAACAGCCAGTCGTCG 148934

Qy      735 --------------------------AlaIlePheSerSerLysGluThrGln------ 743
Db   148935 GCTTCCACTGATGGAGCAACAAGTGGAGCTGGAAGTGCGTCCACAACTACACAAACAGTG 148994

Qy      744 ---------PheAlaGlnIlePheGluProLeuHisLysSerLeuAspProSerAsnLeu 760
Db   148995 CATCCTATTTTCAACGAGATCATCGAGACGCTTCGCCTAAAACTGACACCA---AACTGT 149051

Qy      761 GluHisLeuIleThrProLeuValThrIleGlyHisIleAlaLeuLeuAlaProAspGln 780
Db   149052 GAGCATCAGCGCACAAAGATTGTGACCTTGGGTCACATTGCCTACAATATGCCACAGGCC 149111

Qy      781 PheAlaAlaProTrpLysSerTrpValAlaThrPheIleValLysAspLeuLeuMetAsn 800
Db   149112 TTCCTAACGCCCATTAAGAATATGATTGCGCGACGAATTGTCAAAGAGCTGCTTATCCAA 149171

Qy      801 AspArgLeuProGlyLysLysThrThrLysLeu---------------TrpValProAsp 815
Db   149172 GAA---GTTCCTGCGCAGCGAGACTACGAACTGCCAGAGGACAGTGATTGGTGTGCCCAA 149228

Qy      816 GluGluValSerProGluThrMetValLysIleGlnAlaIleLysMetMetValArgTrp 835
Db   149229 GAGAAACTACCGCCGGACACTCTATGTAAGCTGGATGCGCTCAAGGCTATGGCCAGGTGG 149288

Qy      836 LeuLeuGlyMetLysAsnAsnHisSerLysSerGlyThrSerThrLeuArgLeuLeuThr 855
Db   149289 CTTTTGGGATTGCGTACCGAT---GAGCACGCTGCCCAGAAAACATTCCGAATGCTGGCG 149345

Qy      856 ThrIleLeuHisSerAspGlyAspLeuThrGluGlnGlyLysIleSerLysProAspMet 875
Db   149346 GCGTTTGTTAACCAACGAGGCGATTTGCTCGGCCAGAACCGTCTTTGCGGTGCCGAGAAA 149405

Qy      876 SerArgLeuArgLeuAlaAlaGlySerAlaIleValLysLeuAlaGlnGluProCysTyr 895
Db   149406 TCTTGGCTGCGCCTCGGGGCAGCCTGCGCCATGCTCAAAGTGTGCGAACAAAAGGGCGTA 149465

Qy      896 HisGluIleIleThrLeuGluGlnTyr---GlnLeuCysAlaLeuAlaIleAsnAspGlu 914
Db   149466 GGTGATCAGTACAGCGCTGAGCAGTATTTGCAGCTTTCCCAGCTGATGGTGAGTAGAGAC 149525

Qy      914 ------------------------------------------------------------ 914
Db   149526 ATTCATCTTGTGCTTAGAACTTAGTCTCAATGTCTATCTCGGTGTTCCCGTAGGCTGATC 149585

Qy      915 ---CysTyrGlnValArgGlnValPheAlaGlnLysLeuHisLysGlyLeuSerArgLeu 933
Db   149586 CGGTGCCA-GAAGTTCGGGAAATCTTTGCTCGCAAGCTGCACAAAGGATTAAGCAGAAGT 149644

Qy      934 Arg------------LeuProLeuGluTyrMetAlaIleCysAlaLeuCysAla------ 947
Db   149645 TTGCCCAGGAACTGTTTGCCGCTGGACTTCATGGGCTTGTATGTGCTGGCTGGTCTAGAG 149704

Qy      948 ---------LysAspProValLysGluArgArgAlaHisAlaArg-GlnCysLeuValLy 964
Db   149705 ACTGAGAGGAAGTGAGTAGTTCAAGAAACCATCCGATCCGTTCGGAGAAGTATTGATAAA 149764

Qy      964 s------------------------------------------AsnIleAsnValAr 969
Db   149765 TATTTCAACAGATTGCAAGACCTTGTGCGTCACTATGCAGAAACGGATGTAAACAAACG 149824

Qy      969 gArgGluTyrLeuLys--------------------------------------------- 974
Db   149825 GCGGGAATATCTCAAGACTGTGGCTATGACATGTAGGTGTTTACTTTTGGAATCATATT 149884

Qy      975 ---------------------------------GlnHis-AlaAlaValSerGluLysL 983
Db   149885 GTGTATTTTAACCATTAAATCAAATTATTCCAGCTCCCGACAGCTCAACGGAGTCACAAT 149944

Qy      983 euLeuSerLeuLeuProGluTyrValValProTyrThrIleHisLeuLeuAlaHisAspP 1003
Db   149945 CATTACACATACTACCTGACTACATGCTGGCTTTCGCTATTCCCGTGCTGGTCCACGATC 150004

Qy     1003 roAspTyrValLysValGlnAspIleGluGlnLeuLysAspValLysGluCysLeuTrpP 1023
Db   150005 CACGCTTCACGAATCACGAGGACTACGTACAGCTGCGCAAGATGGAGAAGTGCCTGCGTT 150064

Qy     1023 heValLeuGluIleLeuMetAlaLysAsnGluAsnAsnSerHisAlaPheIleArgLysM 1043
Db   150065 TCATTCTGGAGCCGCTGATGGCCAAACGAGAAACGTTTGTCCATAGCTTCTACAAGCAGC 150124

Qy     1043 etValGluAsnIleLysGlnThrLysAspAlaGlnGlyProAspAspAlaLysMetAsn- 1062
Db   150125 TGCTGCAGCTGATAAAGCATCGCGAGTTCAGTCTGGGG---TCGGACAAGCGCGACAACT 150181

Qy     1062 ------------------------------------------------------------ 1062
Db   150182 ATGTATGTCACTCAATCATTTTCCCAGGCTTTCCCTTCTAAATTATATTTATACATATAT 150241

Qy     1063 -----------GluLysLeuTyrThrValCysAspValAlaMetAsnIleIleMetSerL 1079
Db   150242 TTTTCGGTTATCAGAAAATGTGGGCGCTCTGCGATCTTGCCATGTACATTATCGACTCCA 150301

Qy     1079 ys--------------------SerThrThrTyrSerLeuGluSerProLysAspProV 1092
Db   150302 AGTTCAGTCCATTTGATGGCAACACGAGCACCTTTTCCATGCCGCTG-----------G 150349
```

```
Qy       1092 alLeuProAlaArgPhePheThrGlnProAsp---LysAsnPheSerAsnThrLysAsnT 1111
                ||||||      ::::::   :::|||         ||||||   |||         |
Db     150350 CTTTGCCAGAAATGTATTATAAAGAGCCTGCCGTTGCGAATTTCCAAAACAATGACGTCT 150409

Qy       1111 yrLeuProProGluMetLysSerPhePheThrProGlyLysProLysThrThrAsnValL 1131
              ||:::|||   ::::::         :::|||   |||         |||:::
Db     150410 ATATACCGCTGGACGTG---------TATACGCTGGGAGCCAAATCCACGAGCAAAGCTG 150460

Qy       1131 euGlyAlaValAsnLysProLeu---SerSerAlaGlyLysGlnSerGlnThrLysSerS 1150
                         |||      :::   ||||||:::|||   |||         ::::::|
Db     150461 CCGCAACAGC-AATGACAACGTCGCGAGCAGCAGTGGCTC-CAAAGAGACCGGCGGAACA 150518

Qy       1150 erArg---------------------------MetGluThrValSerAsnAlaSerSerS 1161
              ||:::                           :::   ::::::   |||      :::
Db     150519 GTCAATCATGGATGATGAAAATCCGCAGGTGAGTGAGGTCCTTGGAAAATACTGAGAATG 150578

Qy       1161 erSerAsnProSerSerProGlyArgIleLysGlyArgLeuAspSerSerGluMetAspH 1181
                ||||||                     :::      :::   |||:::
Db     150579 AATCTAACTAGTGTGTGAAATTTGATTTCAGGAGAACAATCTGTTCGACAACATAC---- 150634

Qy       1181 isSerGluAsnGluAspTyrThrMetSerSerProLeuProGlyLysLysSerAspLys 1200
                         |||      |||:::   ||||||   |||   :::::::::   :::
Db     150635 -----------GAGCGGCTGACACTACGGAGCCCATGGCCAAACGAACGCGCGCAGGAG 150682
```

Search completed: September 25, 2003, 00:10:10
Job time : 12397 secs